University of Cape Town

Department of Computer Science

Thesis submitted in partial fulfilment for the award of

Masters
In
Information Technology

Evaluation of the Usability and Usefulness of Automatic Speech Recognition among users in South Africa.

Idowu Modupeola Florence
doju_dupeola@yahoo.co.uk

Supervisor
Dr. Audrey Mbogho.
University of Cape Town

November 18, 2011

DECLARATION

I declare that this thesis is not a copy of any other and every supporting paper has been properly referenced. I declare that this work has not been submitted for any degree elsewhere but to the University of Cape Town for the award of Masters Degree in Information Technology in the Department of Computer science.

Signature: ………………………………………..

Date:

ACKNOWLEDGEMENT

I would like to appreciate the relentless effort of Dr. Audrey Mbogho in making this research study a reality. I appreciate her guardianship throughout this period and her time dedication. Your inputs made a great difference.

To my husband, and the best daughters in the whole wide world, Tunmi & Tomi. I appreciate your support and love all the way!

ABSTRACT

An automatic speech recognition (ASR) system is a software application which recognizes human speech, processes it as input, and displays a text version of the speech as output or uses the input as commands for another application's usage. ASR can either be speaker-dependent or speaker-independent. A speaker-dependent ASR system requires every user to perform training before its usage, while speaker-independent ASR requires no prior training before usage. The technology of ASR is based on identification and comparison of sound patterns; these sound patterns are combinations of the smallest units of sound called phonemes [55]. The phonemes constitute fragments of uttered sounds in speech and their combination gives meaningful sound patterns in languages. There exists a set of phonemes for every language group [51], and associated with each group is the method of pronunciation called the accent. A language group could be identified by the accent in their speech; accent is the set of pronunciation rules of a language group. Accent reflects the cultural divide of a multi cultural society with a common language such as English.

Some commercially available ASR systems are designed based on the accents of the following language groups: English, French, German, Italian, Dutch, and Spanish [15]. These language groups are European with none having any similarities with African languages and accents, (except Afrikaans and English, which, though spoken in Africa, originated from Proto-Indo-European languages [20, 51]).

This study involved the evaluation of commercially available English ASR systems, establishing their usability and usefulness among different language groups in South Africa which use English as a common language. Of particular interest was the effect of African accents on the performance of the ASR systems. ASR technology is widely used and researched in the developed world with reported recognition accuracy of up to 99% [15]. However, English spoken with African accents may have adverse effect on the recognition accuracy.

Despite the fact that most existing ASR systems are not designed for English spoken with South Africans' accents, one can easily purchase them over the shelf in South Africa.

The systems used in this study are:

1. Nuance Dragon NaturallySpeaking, Version10.0 (NDNS).
2. Windows Speech Recognition, Windows Vista version (WSR).

The result of this study indicated that accent has influence on the ASR recognition accuracy. It also indicated that users' satisfaction was greatly affected by the recognition accuracy obtained. The results also indicated poor performance in environments where speech cannot be loud, for example, in the library.

TABLE OF CONTENTS

List of Tables

List of Figures

ABBREVIATIONS:

ASR: Automatic Speech Recognition

WSR: Windows Speech Recognition

NDNS: Nuance Dragon Naturally Speaking

GUI: Graphical User Interface

UCD: User-Centred-Design

PERS: Personal Emergency Response Systems

UEMs: Usability Evaluation Methods

HCI: Human Computer Interaction

IS: Interactive System

UI: User Interface

UE: Usability Engineering

BSI: British Standards Institute

ID: Inclusive Design

ICT: Information and Communication Technology

PC: Personal Computer

SUI: Speech User Interface

CSR: Continuous Speech Recognition

OOV: Out-of vocabulary

SSI: Silent Speech Interface

EMG: Electromyography

ATC: Air Traffic Control

JSF: Joint Strike Fighter

CHAPTER 1: INTRODUCTION

1.1 Background

The history of computing can be traced back to the 1940s. During this era, input devices took the form of thousands of switches and mechanical parts. This era was marked with non-interactive machines (computers) in which users had little or no relevance in the performance or efficiency of the computer because they only controlled switches and toggles [51].

The term ‘users’ became an important aspect of the computer developmental milestones after the first human-computer interface called the graphical user interface (GUI) was developed by Xerox in the early 1970s [51]. This new era created a paradigm shift for scientists in terms of computer technology development, and this era was later enhanced by the introduction of a design technique called user-centred-design (UCD). UCD enabled developers/scientists consider user-centred development of computer interfaces and technologies. This era gave birth to interactive system design. UCD is a design approach which tries to consider how system/computer development could be effective to users or improve users’ performance while considering users’ capabilities and limitations [60].

The GUI is the means of communicating between the users and the computer. It is the means of inputting data for processing into the computer and it is also a means for the computer to display information as output to the users. The GUI was first featured in the Xerox Alto machine in 1973 as an alternative to text-based input. Figure 1 below showed a typical example of a Xerox Alto machine.

Figure 1: Xerox Alto, (Adapted from [84])

The recognition of human speech by a computer (see figure 2 below) is called speech recognition or Automatic Speech Recognition (ASR). ASR is a technology that processes human speech (sound signal) as an input signal to the computer and the computer interprets this sound signal and converts it to text that can be displayed on the GUI or issued as commands [55]. The first glimpse of a GUI with an ASR system capability was pioneered in the work of Fred Jelinek [34]. In 1975, he designed a speaker-dependent ASR system called Tangora, which was a form of Voice-Activated Typewriter (VAT). The Tangora functioned by activating the typewriter with spoken speech, and thereafter converted the spoken speech into sequenced words on a display (GUI) or a paper. The 1970s marked the birth of a new era in which electronically coded human voice complemented the GUI as a new means of input.

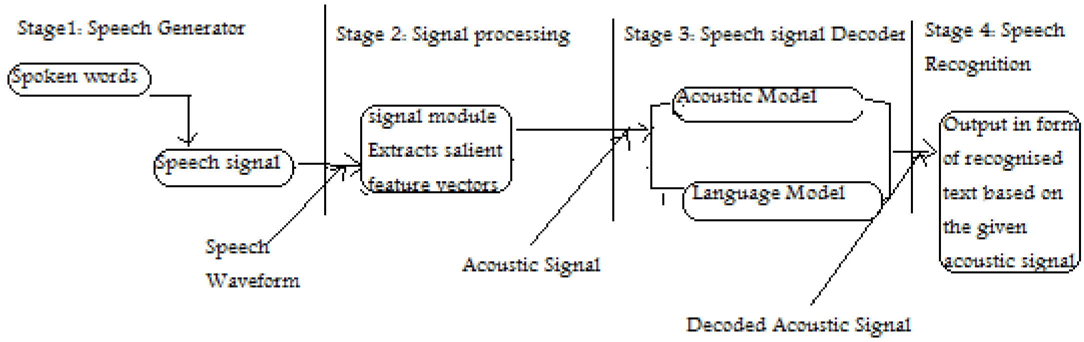

Figure 2: Basic ASR system Operation (Adapted from [55])

1.1.1 Applications of ASR Systems

Juang and Rabiner [34] wrote that the technology of ASR took its place in the history and development of computing because of the continuous effort made by scientists to have a UI which imitates the natural human-human dialogue; they also identified the efforts made in the past decades in order to one day make ASR the user interface of choice. The effort made so far has enabled the use of ASR to appear in many fields such as: medicine, law, journalism, and law enforcement. The use of ASR by the physically challenged for example the blind, paraplegics, and amputees, is also becoming common, enabling individuals who ordinarily would not be able to use the computer carry out tasks similar to those who are physically able [14]. The ASR system acts as a solution to the limitations of the keyboard and mouse for physically challenged users. Furui [21] suggested the need for repetitive tasks (such as information access and transfer) performed by humans to be replaced by an ASR system; this he argued would improve task output by eliminating the distractions caused by hands and eyes of human operators.

Speech recognition can be applied to almost all spheres of human life:

1. Military and National security: The US military was among the first institutes to research into ASR technology and applied it to areas [82] such as the following:

i. Command and control on the move (C20TM),

ii. Soldier’s computer,

iii. Voice control of radios and other auxiliary systems in army helicopters.

Boeing and BAE systems [41] are now incorporating the ASR system into the cockpit of joint strike fighter (JSF) jets; this they say will increase the pilot’s ‘head-out’ time by enhancing an efficient time-share between operating the cockpit by voice and concentrating on the external environment (‘head-out’). In recent times, ASR systems are being used beyond military aircrafts; a Melbourne-based Adacel [41] technology has incorporated the ASR system into training simulators for air traffic control (ATC) trainees; this enables trainees to give commands such as altitude change during training sessions thereby replacing the need to have an instructor present.

2. Medical field: ASR is used in the medical profession to make reports during medical emergencies. It could be used as a way to pre-inform a unit of an occurrence. For example when a patient is being transported from one point to another, ASR serves as a way of sending written reports on the move without the need to engage the hands. ASR can be used while an emergency worker is trying to keep a patient stable. There are ASR systems that dictate directly into Electronic Medical records [15]. Hamil et al [27] reported in their research work the development of a dialogue-based Personal Emergency Response System (PERS) using an ASR system. The objective of their prototype design was to upgrade from an existing emergency push button system used by the elderly living alone, and to create an automated hand-free PERS, which will require speech as a means of request for immediate emergency assistance. Their research work showed ways that ASR systems could be integrated into long-term care facilities for the elderly.

3. Law profession: ASR could also be used in the law profession as a way of converting audio evidences given by a witness into transcribed statements for documentation. It could also be a means by which law enforcement agents can transcribe logs of radio transmissions.

4. Physically challenged users: ASR can serve as a tool to help people with certain disabilities, providing the capability to use a computer system for the physically challenged such as the blind, those with the temporary loss of the use of hands, and amputees [26, 64]. These individuals can use a PC with less aid through the use of ASR systems. Companies can install this system for employees with disabilities in order to give them a chance to work and be productive. The use of ASR in the workplace also helps reduce health hazards such as carpal tunnel syndrome.

5. Telephone-based applications: ASR can be used to enhance operator services in the telecommunication industry. In this case a speaker-independent ASR is used with limited vocabularies but with key words which the ASR system is programmed to recognize from speakers. It is used to save time and money needed to keep employees as operators. Barnard et al [4] wrote in their survey report the possible application of ASR as an integral part of a telephone-based information access system for South Africans; their study identified areas (such as telecommunications) where ASR systems can be applied to benefit users at all levels in a society. Sherwani et al [63] wrote in their report about the design and implementation of a speech interface which was integrated into a telephone-based dialogue system. Their system was designed to search and browse Wikipedia using speech as an input medium.

6. Mobile-based applications: the use of ASR in mobile phones is becoming a popular feature in iPhones and other smartphones. The use of ASR is increasing in these devices; mobile users could easily speak to the phone to check contacts, dial a number and then subsequently speak to someone while driving, requiring no visual attention of the driver hence enhancing safety. ASR systems are now being applied to car navigation systems. In this case, the driver could speak a street address and then the ASR system responds with a synthesized voice the turn-by-turn direction of the requested street. This has been implemented by IBM in partnership with Honda; IBM ViaVoice software has been embedded into the in-built navigation system of Honda car models of year 2005 upward. The navigation system is linked to the street addresses database and the response to spoken request is retrieved from the database and read aloud to the driver without the need for the driver to look on screens or engage the hands elsewhere other than on the steering. This application is available in the USA and Canada [80]. There are commercially available smartphones such as the Nokia E7, Nokia N8, and Blackberry which have facilities for voice commands to control the phones' features.

7. Internet-based applications: ASR could be used to give commands to search the web and perform almost all actions an individual can do using the keyboard to surf the Internet. There are Internet-based ASR applications like Google voice help, which assists users in search via voice input [37].

1.1.2 Usability Evaluation Methods

The ASR system will only be relevant to any user if it will help improve productivity. To ensure that a system is effective and productive to the users, it must be useful and usable. The usefulness or usability is measured by applying usability evaluation methods (UEMs).

An application/system is said to be useful [76] if:

- It enables its user to achieve the set goal or task with ease.
- It improves the performance of the user.

An application is said to be usable if [76]:

- It is easily adaptable for the users.
- It considers its users' limitations and capabilities in its design.

In order to ascertain the usability of an application, usability evaluation is done. Usability evaluation is a technique employed in the discipline of human-computer interaction (HCI), to measure the performance, effectiveness and efficiency of an interactive application, with consideration given to users' limitations and capabilities. Alongside usability evaluation, the usefulness of an application is also established.

There are two approaches [60, 73] to carrying out all known UEMs, these are:

1. **Formative evaluation**: it is a process of evaluating every stage of the developmental cycle of an application before the final design is made available commercially.

2. **Summative evaluation**: it is a process of evaluating a finished product which has completed the developmental cycle of design, and could be classified as finished product, e.g. evaluation of an off-the-shelf product.

This study was carried out using summative evaluation and this approach was used because the ASR applications evaluated are finished products bought off-the-shelf and evaluating them will indicate their suitability for users in South Africa. Table 1 below showed the evaluation approaches and the best possible UEMs to be applied in respective instances. In the case of this study, UEMs such as field observation and the use of questionnaires were adapted from the summative evaluation approach.

Table 1: Evaluation Methods

Evaluation Approaches:	UEMs:
Summative Evaluation	Field Evaluation
	Interview
	Questionnaire
	Usability Testing
	Heuristic Evaluation
Formative Evaluation	Cognitive Walkthrough
	Heuristic Evaluation

1.2 Problem Statement

This research aims at studying the effect accent has on ASR application recognition accuracy, and the usability of the ASR application among African users. The usability evaluation is based on users' satisfaction in terms of the ASR learnability, flexibility and robustness. The immediate goal of this study is to extract a sample of the potential user population and establish usability and usefulness of an ASR system among sample users. The outcome of this study will either show that accent has a negative effect on an ASR application (the recognition accuracy) or show that the ASR under consideration is robust enough to accommodate accent among African users.

1.3 Research Objectives

The core objective of this study is to observe users while using commercially available ASR applications:

1. Nuance Dragon Naturally Speaking (NDNS) Version10.0: this is the DVR Edition with a voice recorder.
2. Windows Speech Recognition (WSR): this is a facility found on recent editions of Microsoft Windows. Windows Vista was used in this study.

To study:

- The relationship between diverse African accents and the performances of the two ASR applications.
- The relationship between users' satisfaction and the recognition accuracy.
- The difference in the word accuracy of the two ASR applications for each user.

To perform the following tasks:

- Training of the two ASR applications for speech adaptation/recognition.
- Voice commands given by users (using WSR).
- Correspondence dictation by users (using both NDNS and WSR).

The users participating in the study will train the NDNS application for 15 minutes as specified by the manufacturer and train WSR for the duration of the training 'wizard' (approximately 6 minutes). Users will perform a task by dictating the same correspondence for the two ASR systems simultaneously, and voice commands will be given to WSR to execute. The environments for this study observation will be based on the type of environment where the target population for this technology would generally use these applications; namely: offices, homes, and classrooms.

1.4 Research Questions and Hypotheses

When the experimenter reviewed the ASR systems that are commercially available, it was gathered that some manufacturers claimed 99% [15] performance for certain accent groups which are outside the accent groups of this study. This study then assumed that since the ASR systems were not designed for its target users, then they are not likely going to experience the same performance. Therefore hypothesis1-I below was set. Hypothesis 1-II was set in order to test the assumptions from online reviews of ASR systems that NDNS has the best performance [16, 17, 18], coupled with the assumption that NDNS would record a better performance because its manufacturer is focussed on the research and development of ASR technology. In comparison, WSR is just one of the numerous facilities of Windows OS. Hypothesis 2-I was set in order to test the fact that user familiarity with a system enhances usage. And it is a common knowledge that most users are familiar with the operations of Windows, On the other hand, using NDNS might pose learnability problems because of user's unfamiliarity with the relatively new package as it is always the case with any new application. Therefore, an observation will help create more insight to the validity of this idea, and the study set to find out the more learnable and flexible of the two ASR systems by setting this hypothesis. Hypotheses 2-II was set in order to qualitatively test the fundamental principles of usability of these two systems. This study is focused on establishing the answers to the following research questions:

1. How well do the two systems under study cope with diverse accents?

Hypotheses 1:

I. The average recognition rate for both NDNS/WSR will be less than 99% for all participants.
II. NDNS will achieve higher recognition accuracy; thus making it more robust compared to WSR.

2. Will users find the two ASR systems learnable, flexible and robust?

Hypotheses 2:

I. Users will find WSR more learnable and flexible compared to NDNS
II. Users' satisfaction will be adversely affected by low speech recognition accuracy.

With regard to Hypothesis1-II, it is important to mention that each of the two systems used its own microphone. While this use of different microphones was not ideal in that it could increase data disparity, there were certain difficulties that made the use of the same microphone or the same audio data impractical. To mitigate this, simultaneous capture of audio data by the two systems was used. These study limitations are discussed further in Section 5.1.2.1

CHAPTER 2: BACKGROUND AND RELATED WORK.

2.1 Usage of ASR as a Speech User Interface (SUI).

The ultimate aim of ASR is to create a natural dialogue with humans, in which the system understands and interprets human speech with appropriate feedback [34].

The early versions of ASR systems were designed to recognize isolated words. As such, users needed to pronounce words with a pause between words in order for the ASR system to recognize effectively. This resulted in an ASR system with limitations in usefulness and usability. The ultimate goal was to design an ASR system that could recognize continuous speech.

The research work in the area of ASR technology has progressed from the time of isolated words to continuous speech recognition. There has been much review on the progress of ASR technology. For example, the work of Juang and Rabiner [34] pointed out the gradual progress of the ASR technology and the eventual design of a system which recognized continuous speech. The review of Juang and Rabiner mentioned the first commercial ASR system (called VIP-100) made by Threshold Technology in the mid 1970s. This was used for quality control in television broadcasting, and by FedEx parcel sorting on conveyor belt. This product according to [34] enhanced further research in ASR technology. In 1987, the former Apple computer CEO, Sculley [62] described a concept called 'Knowledge navigator'. He envisioned a time in which the use of SUI with other multimodal user interfaces (MUI) would form the basis for accessing large networks and databases. His concept suggested the use of ASR technology in a tablet style computer and he proposed hypothetically the commercial usage of this concept for the year 2011. In recent times (as he proposed), this idea can be seen implemented in the latest innovations of Apple such as the iPhones, and the iPads [29, 62]. The guides provided to enhance the design of an interactive system such as those mentioned above is studied in the discipline of human computer interaction.

2.2 Human Computer Interaction

Human Computer Interaction (HCI) is a discipline in computer science which creates guides and means to optimize the interaction between user and computer. HCI is concerned with the efficiency of the system and how it improves the performance of the users. The effectiveness of interaction between the user and the computer is measured as a function of users achieving set goals within an acceptable time frame. This discipline employs tools, methods and techniques in order to ensure that users and computers interact effectively [9, 56, 59]. These tools, methods and techniques are implemented through processes such as usability engineering and usability evaluation. These processes are the means of achieving effective communication between a user and an interactive system. An interactive system (IS) is a system that communicates with the users by receiving information and sending feedback. In practice, HCI activity is carried out through design and evaluation of interactive systems or user interfaces (UI). HCI provides design guides which designers implement during the design and development phases of such IS. These design guides can be categorized into three main principles: learnability, flexibility and robustness. These principles are fundamentals of usability.

2.2.1 Usability

In this study, the focal points are the users and the ASR systems' usability. Usability is an act which encapsulates techniques and tools used in the testing of interactive systems, and in order to fully grasp the concept of usability, it is essential that all that it relates and represents be treated holistically.

Usability means different terms to different authorities but ultimately it is always tailored around the users and their productivities/satisfactions while using an IS. In [1], ISO/IEC9126-1 (2000) defined usability as: "The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". And ISO 9241-11 (1998) defined usability as: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". The IEEE std. 610.12 (1990) also defined usability as: "The ease with which a user can learn to operate, prepare inputs for, and interpret output of a system or component". Nielsen [76] not only defined usability but also identified some of its quality attributes namely: learnability, efficiency, memorability, errors, and satisfaction.

1. **Learnability:** This measures how a new user learns to use an IS quickly and easily and gradually progresses to be an expert user. Learnability is also an HCI design principle which implements users' core psychological properties in the design lifecycle, [49, 53]. The ISO/IEC 9126-1 (2000), defines usability in terms of learnability [1]. Learnability comprises five components [39, 40, 49] which are important to the overall effective usage of an IS. These are:

    - Predictability: Designers need to consider users' experience with other similar system in the implementation of a new system. Users can apply their prior knowledge to interact with and predict the activities of a new system. The ability of a new system being predictable to a user results in an effective interaction.

    - Synthesizability: This guides the designer to ensure that communication between users and IS is transparent; users are informed of status of information exchange, thereby building effective communication with feedback. The user should find the interaction status observable and measureable in the form of updates or warnings. These visible updates will effectively reassure users that the system is still responsive.

    - Familiarity: This is aimed at guiding designers to develop a system which the users find easy to interact with based on past knowledge of interaction. Users want to use past experience with other systems to navigate through any other similar systems. It is therefore of paramount importance that a system designer represents iconic features such that there is visual recognition with the users. This will boost the performance of the system and users' experience will be positively impacted.

- Generazability: This design guide is concerned with user ability to apply knowledge in an area to other areas. For example the copy/cut terms could be used over a wide range of systems and still retain the copy/cut concept. This guide positively enhances users' experiences by preventing the need to learn new terms each time a new system is encountered.

- Consistency: It is important that a system maintains consistency in all its operation. A user navigating through a system would find an interface difficult to use if an icon represented more than one concept. This could occur due to error in design and therefore the need for evaluation at the early stages of the design phase.

2. **Efficiency:** The moment users have established ability to use the system, the next focus is on the effectiveness of the system in boosting their performance. This is measured over continuous usage of the system.

3. **Memorability:** The user's ability to recollect the process of operation of a system after a long absence of usage is a factor of great importance in usability. This will reduce the need for continuous re-training of personnel.

4. **Errors:** This refers to the rate of occurrence of error, and it could also be used as a measure for usability; it plays an important role in usability evaluation. The ease of recovery from an error is also important to the overall usability of a system.

5. **Satisfaction:** Users' satisfaction/ experience will influence their performance and willingness to continue using the system.

In the practise of HCI, usability is defined in terms of implementation of the three main design principles**: learnability**, **flexibility**, and **robustness**. Since learnability has been discussed extensively above, focus now turns to flexibility and robustness.

Flexibility is a design principle which involves the ease of flow of interaction between the user and the system. This principle is concerned with ensuring that all ambiguity in information flow is eliminated [59]. To fully grasp the concept of flexibility, it is worth mentioning that its application involves five components [39]:

- Dialogue Initiative: This is when the flow and exchange of information is controlled by either the user or the system; it is called system pre-emptive when the system controls the rate and flow of information, while user-pre-emptive is when the control of flow of information is solely by the user. During start-up, system pre-emptive control is applied and this enables the users to utilize the system.

- Multithreading: This is when a system allows a user to perform multiple actions; multithreading can either be concurrent or interleaved. Concurrent is when more than one action can be performed simultaneously while interleaved is when actions are performed one at a time on a particular application.

- Task Migratability: This is when system and user interaction is such that action execution is transferred between user and system. This is to provide a means of easing multiple tasks off the user, and providing the user with the opportunity of optimal performance.

- Subtitutivity: This provides users with options on how to input and output information.

- Customizability: This is providing users with freedom to adjust the system to meet individual preferences.

Robustness: It is a design principle that is concerned with the ability of the system to respond to changes and diversity and still perform optimally. It relates to the ease of interaction between the user and the system. Systems that implement this principle give users sufficient information on task completion, ability to identify and recover from errors [60].

These design guides mentioned so far highlight the importance of considering a system design from the viewpoint of human perception and nature. Furui [21] identified the problems of ASR systems as being those of robustness to: (i) speech variations, (ii) adaptation/normalization to changes in environmental conditions, and (iii) ease of human-machine interaction. His work recognized the importance of robustness in ASR system design.

Abran et al [1] defined usability in terms of users and identified three sub-groups of users; in their own view, the meaning of usability is distinct to each user group as illustrated in table 2 below.

Table 2: User-Subgroup and Usability Definition [1]

User- Subgroup:	Usability Definition:
End-users:	Ability of the system to enable the user achieve optimal performance with efficiency and time optimization.
Manager:	Ability to make informed decision on choice of system that will best suit the need of the organization and ultimately improve staff overall performance.
Software developers/designer:	Usability means ability to develop system internally with focus on design, documentation and maintainability.

2.2.2 Relevance of usability

Users are naturally and consciously seeking a usable system; these systems could be electronics, household appliances, mobile phones, or websites. The reaction of a user to a difficult system will depend solely on the availability of alternative systems. In the case of a website, the importance of usability cannot be underestimated as users are only a ‘click’ away from alternatives. According to Nielsen [76], usability is the key to survival of a website. He pointed out that a website that lacks the usability attributes sees its users decline as there are other websites that render similar services with a more learnable and efficient interface. In the case of Intranet, staff productivity is affected according to the usability capability of the systems in place. On ecommerce, the less time a user spends trying to purchase an item, the shorter the queue on the ecommerce server. This translates to quick and smooth running of businesses, and ultimately more financial returns.

2.2.3 Usability Evaluation.

For the ASR systems under investigation in this study, the only means of verifying their usability quality attributes discussed so far is by the use of evaluation processes. Evaluation can enhance development if it is part of the design and development lifecycle [60], as problems identified during an evaluation process result in developers improving subsequent versions of the system for better performance [30, 34, 42, 59]. Evaluation ascertains the capabilities of a system in terms of its usefulness and usability. Monk [46] and Karat & Karat [35] indicated in their works the relevance of users in usability research. More companies in their marketing research strategies are now adopting user involvement and evaluation in their product development processes [66]. Evaluation provides a platform for performance verification of human computer interaction.

Usability evaluation can be applied to an IS in four basic ways [50, 74, 75]:

- Automatic method: using programs to measure usability specifications
- Empirical method: this involves usability testing with real users.
- Informal method: Evaluation based on evaluator’s experience only.
- Formal method: usability specifications measured by the use of formulas and models.

The empirical method is the most common because of its ease of use. It adopts either user testing or usability inspection methods. User testing and usability inspection involve users and evaluation experts respectively. User testing is not in the absence of evaluation experts but rather users interact with the IS while experts observe and interpret results. Usability inspection is done without direct involvement of the users; it could be in the beginning or in the middle of the design/development phase of the IS. For each of these methods, there are UEMs that are applicable as shown in table 3 [50].

Table 3: Comparison of User testing and Usability inspection methods

User Testing-UEMs:	Usability Inspection UEMs:	UEMs Descriptions:	Advantages	Disadvantages
Field Observatio-		Evaluation of users in their	-Users are in their natural environment,	The influence of noise could

n		natural environment.	thus, result obtained is free from psychological effect of a controlled environment.	adversely affect user's performance.
Interview		This is one-on-one dialogue in which observer (interviewer) questions the subject/participants.	-It is flexible to change and gathers different users' opinion. -There is opportunity to fully explore an area of interest to either subject or interviewer. - More unanticipated problems could be identified from information gathered.	-It is subjective. - It gathers unstructured information. -information gathered is difficult to interpret
Questionnaires		A structured list of questions used to gather information on users requirements/ expectations of a system.	-Quick to use - Can reach as much audience as possible simultaneously. -Data collected is easy to analyse	- It is limiting regarding users opinion - It is subjective
Usability Testing		Observing users in a controlled environment	Users are given clear tasks to perform. -There is no distraction -performance is easy to measure	User's behaviour could be altered due to the controlled environment. -thinking aloud and performing tasks could be tedious to users.

	Heuristic Evaluation	Experts carry out individual inspection on a system in order to identify problems with the user interface [48].	-Collaboration of Independent results help make informed decisions on interface problems.	- Requires more than one expert. -problems solved do not necessarily guarantee users' satisfaction.
	Cognitive Walkthrough	Experts apply a set of steps to ascertain that the interface will deliver on learnability [43].	Helpful when the interface is in design/development phase.	It is liable to miss users' expectations.
	Pluralistic Walkthrough	Experts, users and designers gather to identify problems relating to a design [6]	All stakeholders are represented, thus, more likely to produce a usable interface	Clash of interest could lead to delay in design
	Standard Inspection	System compliance check by an expert.	Systems not meeting up with standards are easily identified	System failing the compliance check will lead to re-work or unusable product.

Nielsen [49] argued that: "designers are not users and users are not designers". Abran et al [1] also identified the different users and their expectations on usability. This necessitates the need for evaluation by independent evaluation experts, whose role is to bridge the gap between users and developers/designers' usability expectations. From Table 3, there is no UEM which is perfect and sufficient to measure usability. Nielsen [50] advocated for the combination of more than one of these UEMs in any usability evaluation process in order to achieve more valid usability evaluation results.

2.2.3.1 Usability Engineering

While usability evaluation provides the means for measuring the performance, effectiveness and efficiency of an IS, it does not provide the method for achieving the usability attributes. Usability Engineering (UE) is the structured process which provides methods for achieving usability of an IS. Throughout the process of UE, different UEMs are required in order to develop a usable IS. These UEMs are applied to the design process by the use of sources such as: target population (users) for usability testing /user testing or the use of usability experts to carry out usability inspections. Usability testing is done when users apply the IS to carry out tasks; the performance achieved is then used to estimate how the application would support the user to complete a task. Usability inspection is done when experts (developers, evaluation experts) analyse the usability aspect of an interface. Whatever method is implemented, the ultimate goal is to achieve usability of an IS. The process of UE is divided into three phases:

1. Requirement Analysis
2. Design/Testing/Development
3. Installation

1. Requirement Analysis: The requirement analysis is the phase in which the usability goals and functionalities of the proposed IS are defined. This phase marks the beginning of the developmental cycle of an IS. This is when the target population for the IS are established and the functionality explicitly defined with design guides embedded in the set goals. The developer/designer of the IS has a blueprint of the intended product at this stage and would carry on to the next phase after fully understanding the usability goals and expectations of the IS design [49, 59]. During requirement analysis phase, when a more universal product is desirable for a heterogeneous group of users, a more UCD approach such as inclusive design (ID) is adopted [33]. The British Standards Institute (BS7000-6) [33] defined inclusive design as: "The design of mainstream products and/or services that are accessible to, and usable by, as many people as reasonably possible on a global basis, in a wide variety of situations and to the greatest extent possible without the need for special adaptation or specialized design." Inclusive design adapts a proactive approach to requirement analysis by considering in addition to target audience, audience of different geographical locations, age, race, and with different capabilities and limitations. It considers the widest possible audience in order to create a universal usable design. Keates and Clarkson [37] modelled a five level design approach (as shown in figure 3 below) to inclusive design, which could be used in formulating the requirements of the first phase of UE.

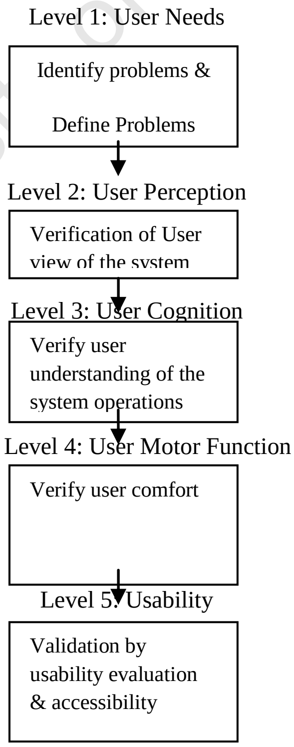

Figure 3: Model of Inclusive Design (Adapted from [37])

Keates and Clarkson reported in their work a successful implementation of this model in the design of a software interface for the control of an assistive technology robot [12, 37]. Level 5 of this design model needs to be part of early design and development for there to be a significant impact on design and at the best cost possible. When usability evaluation process/result is applied to the early design phase, it results in a cost effective change, and makes the best possible impact on design. Tomlin [69] illustrated this in his graph (see figure 4 below) on how early changes to design could make best possible impact. He argued that the best result of usability is achieved at the early stages of design and development when changes made are at a minimal cost.

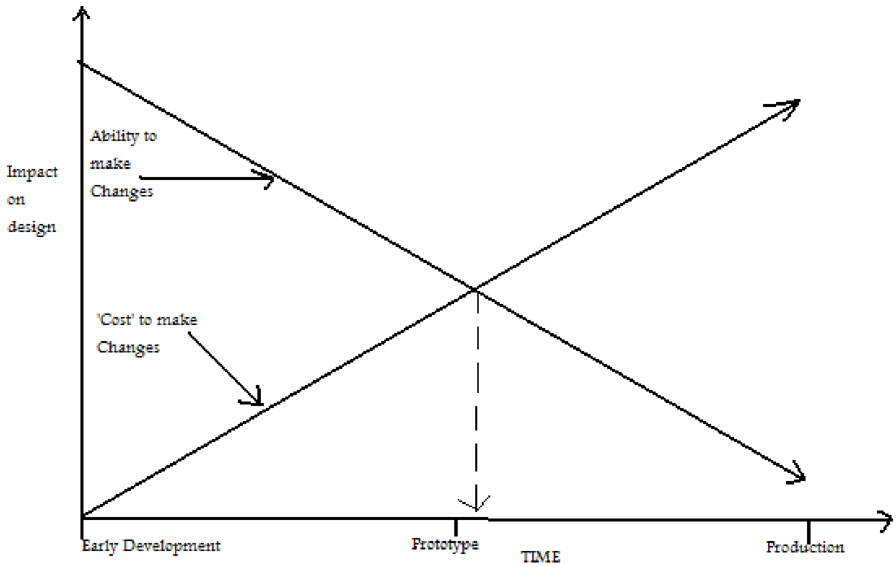

Figure 4: Time Vs. impact on design (Adapted from [69]

2. Design/Testing/Development: After the requirement analysis has been fully stated and documented, the second phase is then implemented. During implementation of the design/testing/development phase, a formative evaluation is done, which results in iterative testing of the initial design. Each time a formative evaluation is done, corrective measures are then reapplied to the design in order to achieve the set usability goals. Iteration of formative evaluation enables designers receive feedback through user testing and identify potential problems on the initial, prototype and then to the final design [49, 59]. During this phase, one of the UEMs such as cognitive walkthrough is applied by usability experts who inspect from the initial to prototype design, to establish learnability before the final design is implemented. Heuristic evaluation is also applied in order to ensure flexibility and robustness of the IS. Formative evaluation is done in order to achieve the target functionalities of the application/product.

3. Installation: In this phase, the final design of UE processes is validated by the use of summative evaluation (applying appropriate UEMs). Summative evaluation is done on the final design in order to establish that it will deliver on its usability goals. This UE phase involves a formal evaluation and documentation of the final design efficiency, effectiveness, and performance. Summative evaluation is the final usability evaluation approach applied on the finished design before it is released for commercial purposes. Usability testing is the UEM applied to evaluate the final design. The result of this evaluation will inform the designer on the validity of usability goals and design functionalities. The installation phase marks the end of design/development phase of the UE process [49, 59].

After the installation phase of the UE process, the success of the new product in the market place is of great concern to all stakeholders in the design. Hanna et al [28] described the success or failure of a system as being dependent on effort applied to fully study and evaluate its target population. Their arguments are in line with the expectations of the usability experts. For the usability experts, a system is classified a success if:

- It improves the performance of its users.
- Users do not feel insufficient in knowledge while using it because of unfamiliar terminologies.
- Users find it easy to learn
- Users can easily relate its contents with a recognized perception.
- The system is robust enough to accommodate diverse users.

On the contrary, success to investors in the system design is measured by monetary gain. Therefore, there is always a two-way success measure in any design. It is worth noting that the success of a system in terms of usefulness and usability does not necessarily translate to success in the market place. Factors such as:

1. Presence of other competitive and well branded products.
2. Poor marketing strategy
3. Digital divide

could make a system which has been properly designed fail to reach its maximum performance potentials. Despite ensuring a system's usability and usefulness, it is important that the system gets to the target population; this can be achieved by ensuring that supply is made available in the geographical location of the population with a good marketing strategy put in place. All necessary steps should be taken to ensure that users know the difference between what other competitive systems are offering compared to a well designed system; this can also be achieved by stating clearly features of the systems. Factors such as the digital divide could have a great negative impact on a product's success. Digital divide is the extent to which different groups of people have access to, benefit from, or experience enhanced performance with regard to information and communication technology (ICT) or the internet. This divide may be as a result of race, economic status, gender, or location. Global digital divide refers to the extent to which different countries have access to ICT and the internet [13]. Nielsen [13] staged the digital divide into 3 groups: (1) Economic divide, (2) Usability divide, and (3) Empowerment divide; of importance to this discourse is the usability divide, which he argued posed a great threat to the benefits and success of any technology.

2.3.1 ASR Performance and Constraints on Usability.

An ASR system is a form of speech user interface (SUI) of an IS and in its design and specifications there are known constraints which affect its performance. Some of these constraints are pertinent for improved performance of the ASR, while other constraints are due to limitations in knowledge of the ASR technology. The different ASR systems designed thus far have with them different limitations on performance and the review of these different ASR systems has informed on the choice of ASR systems used for this study. The choice of ASR systems for this study was based on:

- Best possible performance for all known ASR systems.
- Affordability to the target population.
- Commercial availability to the target population

- Best possible performance for all known ASR systems: This was based on reviews of commercially available ASR systems: IBM ViaVoice, NDNS, Voice Express, Nuance Macspeech (for Mac OS) Dictate, e-Speaking, TalkingDesktop, Windows WSR, Phillips SpeechMagic, SpeechWorks and Voice Finger. NDNS was rated the leading ASR with best possible performance for all existing ASR system. These reviews were gathered from Internet sources [16, 17, 18, 44, 67, 70, 71, 72, 77, 78, 81]. NDNS (for Windows) was preferred over Nuance Macspeech (for Mac OS) as the review of the target population indicated the popular usage of Windows compatible PCs. Table 4 below illustrates the list of best speech recognition software according to [44]:

Table 4: ASR System rating

ASR Application Software	System Rating
Dragon NaturallySpeaking (NDNS)	1
Microsoft WSR	2
IBM ViaVoice	3
MacSpeech	4
Philips	5
SpeechWorks	6

Tellme Networks	7
Julius	8
CMU Sphinx	9
CSLU Toolkit	10
HTK	11
Voxforge	12

While there are open source ASR systems such as Sphinx, Julius, Simon, Atros, and Tazti which are free to download, they are constantly undergoing design changes and research reviews. These open source ASR systems are mainly for research and development [2, 3, 10] and therefore not the focus of this study. WSR was chosen because of its ease of acquisition; since Windows Vista (and later versions) includes speech recognition capabilities. This implies that by either upgrading or purchasing the latest versions of Windows compatible PCs, a user automatically has access to ASR application.

- Affordability to the target population: there are different versions of the NDNS systems sold at different prices [52]:

1. Nuance Dragon Naturally Speaking 11 legal- costs R 7,855.86.
2. Nuance Dragon Naturally Speaking 11 Professional 11- costs R 5,144.99
3. Nuance Dragon Naturally Speaking Premium 11.0 (5 users) - costs R 6,180.95
4. Nuance Dragon Naturally Speaking Professional wireless- costs R7, 250.00
5. Nuance Dragon Naturally Speaking (NDNS) (Version10.0) ASR, DVR Edition (with Philips voice Tracer and Recorder) – costs R 3000.00.

The above lists of Nuance products were selected for consideration, as they are all designed for professionals. Product 1 was not used because its target population is mainly those in the law profession and since we are considering other professionals, this product was ruled out for possible evaluation. Products 2, 3, and 4 are general software that could be used by any individual from any discipline but price indications of these ASR systems are expensive compared to Product 5, which serves the same purpose but at a reduced cost. Therefore Product 5 was chosen over Products 2, 3, and 4 because of its affordability

The ASR systems have evolved over time in terms of their performance and efficiency. There are ASR systems which claim 90-99% recognition accuracy [15, 42]. Despite the performance claims such systems are still faced with constraints [42]. In the case of the WSR facility, it informs users beforehand that performance would be adversely affected if the accent is not British or American

[45]. Both the speaker-independent and speaker-dependent ASR systems still encounter constraints which consequently affect their usability and usefulness [42]. These constraints are:

1. Spoken Words: The earlier versions of ASR systems were isolated word recognition systems, and later developments resulted in continuous speech recognition (CSR) systems. The ASR systems used in this study are CSR systems. However, the commands for UI use isolated word recognition, which makes WSR both isolated and CSR application. It is worth noting that the users' speech is influenced by factors such as:

1.1. **Co-articulation**: this is an occurrence in which words spoken in a continuous speech overlap, the final sounds of a word is lost to the sound of the beginning of the adjacent word thereby creating a new word/phrase and meaning. In this process, meanings are lost and communication interrupted because the final parts of phonemes of uttered sound are assimilated into the next uttered sound, thereby changing the word spoken into a new phrase or word. This is one of the factors affecting ASR systems' usability and usefulness [7, 38, 57]. This factor is influenced by:

1.1.1. **Speaking rate-** the speed with which an individual pronounces words; this pronunciation could be in a slow, steady form or could be a rapid pronunciation. When continuous speech is spoken slowly, the words of such speech would appear as if they were isolated words and no part of the phonemes will be lost. On the contrary when continuous speech is spoken rapidly, some of the phonemes are lost in translation [57]. The example of [57] showed the effect speaking rate of continuous speech has on phonemes and consequently causing loss of phonemes. When phonemes are lost, words lose meaning and this results in recognition error. ASR design and development is still challenged with this human factor and as such the ever present variation in recognition rate for users affects usability.

1.2 **Spontaneous speech** is affected by:

1.2.1 **Disfluencies**: these occur in unstructured, unrehearsed, unprepared, and unedited speech uttered without any correction. Spontaneous speech is characterized by word repetition, hesitations, and filled pauses. Despite the ability of ASR systems to detect and correct disfluencies, recognition rate is still adversely affected [36]. The language model structure (LMS) which is the language lexicon of any ASR in which identification, comparison, and recognition of words takes place is affected (consequently affecting recognition rate) because:

(a) Time is wasted on substitution and deletion of filled pauses and word repetitions.

(b) Disfluencies increase deletion, insertion, and substitution error because the LMS of ASR systems is designed to recognize syntactically and semantically structured grammar. Consequently, speech variation from this structure leads to error [22, 25, 36]. The work done in [36] indicated that signal processing algorithms designed to detect and correct disfluencies before transcription increased NDNS recognition rate. This work indicated that ASR systems still lack the capability to completely eliminate human factors adversely affecting recognition rate.

1.2.2 **Out-of vocabulary (OOV) words:** OOV are words that cannot be found in the language lexicon of ASR systems. This is a factor affecting ASR performance. Though most recent ASR systems are designed with large vocabulary they still have limited vocabulary size or closed vocabulary [85]. A closed vocabulary is necessary because of the run-time requirements and limited word training in the LMS [85]. When an ASR system encounters an OOV word, this sometimes leads to recognition error, as the system tries to replace such a word with a similar sounding word. As a result of this constraint, every ASR system has different means of detecting and adjusting to OOV words. In [15], the features of NDNS inform that it can easily customize OOV words and add them to its vocabulary but it is possible for similar sounding words to result in recognition error. An ASR system tends to have more recognition error as it customizes OOV words into its LMS, as language structure is altered and new phrases are produced from the language lexicon and OOV words.

2. Noise robustness: the design and development of ASR systems often use 'clean speech' [11] as the source of speech signal, and this design and development is carried out in a noise-controlled environment. On the contrary, the human environment in which these systems are used is characterized by additive noise [79] (ambient noise, mismatched channel and active interference signal). It is worth noting that most systems undergoing evaluation by manufacturers record high performance because most of the evaluation process is under controlled environment [11]. The real test of usability arises when these systems are in the hands of end-users. Users' natural environments range from the office to home and many more. The ambient noise of these environments varies from place to place; a user who works at the airport might not be able to achieve high performance while at work because of the presence of interfering noise. Despite recent ASR systems having high robustness to noise, there are instances as mentioned above in which noise inevitably reduces performance. The robustness to noise can be achieved by capturing the speech signal before its exposure to ambient noise and this can be achieved by the use of noise-cancelling microphones. These microphones are also prone to convolutive noise. Convolutive noise occurs when the impulse activities of the speaker interfere with the output of the microphone [79]. While avoiding ambient noise, there is still the convolutive noise impeding performance. Noise is still a human factor which adversely affects performance of ASR systems.

3. Speaker-dependency: the system in this case has been trained to recognize speech with a set of rules in place such that words are recognized based on the pronunciation the system has been trained to accept. Any variation in the speech could lead to the system failing to recognize a word in its vocabulary. This is an issue of users' accent; because ASR systems have been designed using a particular accent group, a deviation from this accent results in poor recognition rate. Accent indicates the rules of pronunciation of phonemes, and differs from one language group to another and it is as a result of:

1. Deviation from a standard language which is known as the general language: this happens when an individual imperfectly learns the pronunciation rules of the language; this eventually creates phonemic differences, and many other linguistic differences as discussed in [31].

2. The influence of mother tongue could create interference which results in phonological rules of mother tongue being transferred into a learned language such as English [31].

3. The age of learning can also influence the accent in that learning a language at a younger age will result in little or no difference of accent from that of the learned language, while learning at an older age will result in an "in-between" accent [31].

4. Migration also causes accent to change from one of mother tongue to an accent though similar to the new language but still influenced by mother tongue. In [24], Gough mentioned how South African diverse English accents resulted from migration. He also pointed out how accent reflected the social and economic divide in South Africa.

5. Pronunciation rules: There are different accents in the English language, and they can be classified as rhotic or non- rhotic. Rhotic involves the pronunciation of the phoneme |r| in any given word irrespective of its position in the word; for example stork. Non-rhotic is the voiceless pronunciation of the phoneme |r|, except when it is followed by a vowel, for example, rain, run and rug. South African English belong to the non-rhotic accent group [19]

These human factors mentioned so far are human limitations, and as usability involves consideration of the users' limitations, the design and development of ASR systems need to fully consider these human factors.

In this study, the effect of speaker-dependency on usability was considered, and the effects of the other human factors were minimised by:

- Encouraging the users to speak steadily as a news caster.
- Ensuring that the environment used was free from ambient noise.
- Ensuring that the microphone was properly placed.

The aforementioned conditions were required in order to establish the relationship between accent and recognition rate without the influence of these other human factors.

2.4 The Current State of Research in ASR system Design and Development.

Existing work on ASR applications dates back to the 1950s. During this era, ASR systems worked best for simple words [34] and researchers were faced with the following challenges [42]:

1. Cost
2. Real-time response
3. Speaker dependence
4. Robustness to variation such as noise, microphone, speech speed and loudness.

5. Inability to recognize continuous speech

From the 1960s upward, ASR systems have evolved from single words recognition systems to those with the capability of processing continuous and complex speech [34].

In the 21st century, research work is being done to provide the ultimate ASR systems with the capability of understanding speech signal as in human-human communication [11]. This is the present state of research in ASR design and development. Existing research work on ASR systems is focused on developing systems that are robust to the human factors mentioned above. The work of [61] indicated effort to overcome the effect of ambient noise on ASR systems. This research work aimed at improving the performance of existing ASR systems. It is an established fact that for an individual to use an ASR system, the uttered speech must be audible. Citing a scenario in which an establishment decides to adopt the use of ASR systems, an office environment could be turned into a noisy environment of chattering staff. And the use of ASR systems in services which require confidentiality and privacy could be compromised because of the need to utter audible sound. The research work of [61] proposed a silent speech interface (SSI) in which the proposed SSI allows users to work with IS with inaudible speech. The research on SSI considered the use of Electromyography (EMG) – this involves interpreting users intentions based on muscle activities. Another research work worth mentioning is [36] whose focus was to eliminate disfluencies by introducing a standalone signal processing algorithm; this algorithm was used with a commercially available ASR system (NDNS). The result of this research indicated improved performance of NDNS after the signal processing algorithm was applied to the speech signal. Also the research of [85] proposed a method for detection of OOV words and also to create a method for open vocabulary.

Currently, ASR systems are becoming more and more commercially available to users. One such system is Google Voice Search for mobile devices, which allows users to verbally present queries to the search engine. It carries 25% of USA Google mobile search [37]. This application supports popular phones like: Android, Blackberry, iPhones, and Nokia S60.

2.5 Related Work.

The research work of Doe [14] measured usability among users whose first language is English. He measured accuracy as a function of error correction time and studied the relationship between training time and accuracy rate. His work focussed on establishing the effect training time, error correction time, and different correspondence types (business, personal and technical) have on the accuracy of an ASR system. The participants of his study are different from those of this study but informed on accuracy rates of NDNS among research participants.

Barnard et al [4] surveyed the viability of an ASR for South Africa's eleven official languages. The research focussed on the technical issues that needed to be addressed before a telephone-based ASR information access system could be a reality in South Africa. It reviewed the technicalities of ASR technology. The focus was to develop a dialogue system in which an ASR facility is embedded in a telephone-based platform for the purpose of information access. The work considered the same target population as this study but differed as it carried out a requirements analysis.

Hugunin and Zue [32] compared the performance of an ASR system to that of the conventional keyboard/mouse. This study focussed on system performance by evaluating and comparing two different input media. Their work helped gain more insight into the capabilities of the ASR systems when compared to the conventional keyboard/mouse. Their work differed from this study as its focal point was on system performance and not the user.

The study in [30] established the need to involve the target population in the design and development of any system with its focus on speech recognition integrated into telephone services. It pointed out the relevance of qualitative and quantitative methods of usability evaluation in emerging technology. This research work considered ASR for Finnish users using a telephone-based platform. This work is similar to this study as it did consider users but differed because it used Finnish users and a telephone-based ASR system. It motivated this study to proceed and explore ASR performance among users in South Africa.

The work of [53] was designed around an existing information system (radio broadcasting). A voice interface was designed for rural Indian farmers, and user testing was also performed on the design. It reported challenges faced during the research and gave helpful hints to readers interested in designing similar system in developing regions of the world. The design provided a means by which farmers could access and contribute vital information (for example weather reports) that is relevant to their occupation. This work could provide helpful hints in the future work of ASR system development for South African users. It also informed that the financial commitment need not be too expensive.

In [63], a system called Voicepedia was developed. This design integrated ASR into a telephone-based platform to search Wikipedia. This work also compared the performance of Voicepedia to Smartpedia (a smartphone GUI-based alternative) by performing a search on Wikipedia using both. The study aimed at providing means for non-literates to access information on a platform that they are familiar with. It also informed on the expanse of ASR application and it encouraged the need for this study to explore the possible context of use among South African users.

Sirpa [64] presented in his work the experiences gathered from seventy-two usability evaluation processes. This work aimed at informing on the usability methods that best suit systems at different phases of development process. Sirpa accomplished this by reviewing the experiences gathered from seventy-two evaluation processes that he performed. The results obtained were based on the accumulated results of these evaluations and as such made inferences on what method best suits a development phase. The evaluation review of his report was informative in the planning process of this study.

2.6 South African Users and ASR systems.

South Africa is a multi-cultural society with eleven official languages including the English language and Afrikaans both of which originated from Europe. The other nine languages are indigenous South African languages. English language in South Africa is the language of cross-cultural communication in institutes of learning, offices and government establishments, where also interactive systems are found. According to the Laschinger and Goldstruck [58] report on South Africa PC users statistics, PCs in use in South Africa reached the five million mark as of July 2006. In their study, they reported that the estimated life span of desktop PCs ranges between 3 and 6 years and that of laptop is at most 3 years. They argued that desktop PCs are likely to remain in usage as second hand PCs after software upgrade but that laptops are more likely to be replaced after their life span. In Goldstruck's words [58]: "Laptop computers cannot be upgraded as easily or as cheaply as desktop PCs, so they have a shorter useful life". From this study report, one can deduce that as these 5 million users (as of 2006) upgrade their desktop PCs or replace their laptops according to the life span estimates, a substantial number out of this 5 million South Africans (not including new owners) would unwittingly have acquired the ASR facilities in their desktop PCs or laptops as Windows Vista and later versions of the Windows operating system come equipped with speech recognition capabilities. Despite this possible large number of users with ASR capabilities, no research records or reviews of WSR user testing in South Africa was encountered during this study. PC users are unwittingly absorbing the ASR technology into their daily life; therefore it is worth the effort of this study to observe how sample population will fare using this new technology.

NDNS is sold off the shelf in South Africa. Its features include:

1. Document creation.
2. Ability to create and send email.
3. Searching the web by voice command

The web site for NDNS indicates that the software is designed for users such as students, professionals and the physically challenged [15]. Similarly, a sample of South African students and professionals are chosen for participation in this study. NDNS describes its features and capabilities but does not include any warnings regarding the negative effect of accents on performance. This study seeks to provide an objective, independent assessment of both NDNS and WSR and establish how suitable they are among diverse South African English accents.

CHAPTER 3: METHODOLOGY

3.0 Study Methodology.

In the case of ASR systems this study is reviewing, despite recording a success rate of up to 99% [15], there are no established facts or claims by the manufacturers that African users will achieve the same success rate. There are no means to verify that its UE process involved African users in its design and testing. Windows compatible PCs are popular and well branded products, and adding speech recognition to the application package could be seen as an extra feature for better performance, but for African users, would this mean a better performance? The different participants used in this study will shed more light on any possible existence of a usability divide with ASR system target population.

This study is interested in the evaluation of two ASR systems which have been primarily designed and developed for users in America and Europe but also sold in some regions of Africa, such as South Africa. This study result will give an insight to any existence of a digital divide (with regard to usability).

The research questions for this study are:

1. How well do the two systems under study cope with diverse accents?
2. Will users find the ASR systems learnable, flexible and robust?

The UEMs of interest to this study are: field observation and the use of questionnaires. The field observations entail observing users and will be used to answer the first research question. And the use of questionnaires will be used to answer the second research question. The field evaluation is concerned with users testing the two ASR systems of interest; this process will establish the usability capabilities of the ASR systems. The use of questionnaires will indicate users' satisfaction and also inform on users' opinion on learnability, flexibility and robustness of the ASR systems. Learnability will be evaluated in the questionnaire by finding out how easy users find training and using the ASR systems for the first time. Flexibility will be evaluated by finding out the ease with which users navigate and communicate with ASR systems. Robustness will be measured based on average performance of each ASR system for all the users who participated.

3.1 Participants.

Selection criteria were based on the most likely individual that can possess an ASR system, or are unwittingly in possession of the ASR system. The following professionals were identified: doctors, nurses, writers, entrepreneurs, and students. These sets of participants took part in the study after they were approached and they gave their consent for participation. Forty participants were approached, but only twenty participants eventually made time out to participate. There were eight male participants and twelve female participants. Participants were asked to fill a questionnaire in order to:

- Ensure that they are computer literate.
- Ensure their ability to speak the English language.
- Ensure that they are literate enough to follow the study procedures.

3.1.1 Pilot Test.

A pilot test was conducted; using only NDNS, during which three participants were used; two with African accents and the third with a British accent. There was a great variation in the speech recognition rate: for the two African accents, performance was poor and for the British accent performance was good. The metric used to assess recognition accuracy is the word recognition rate (WRR) and it is the percentage of words that are correctly recognized; the higher WRR the better the performance. Table 5 below shows the result of the three participants. The graph (see figure 5 below) indicated the WRR and this study adopted the use of WRR for the actual result analysis.

Table 5: Comparison of Subjects Word recognition rate

Subject 1: African accent	Subject 2: African accent	Subject 3: British accent
Total words dictated: 248 Number of words correctly recognized: 41 words Number of words missed: 207 words WRR =16.5%	Total words dictated: 248 Number of words correctly recognized: 71 words Number of words missed: 177words WRR=28.6%	Total words dictated: 248 Number of words correctly recognized: 223 words Number of words missed: 25 words WRR=89.9%

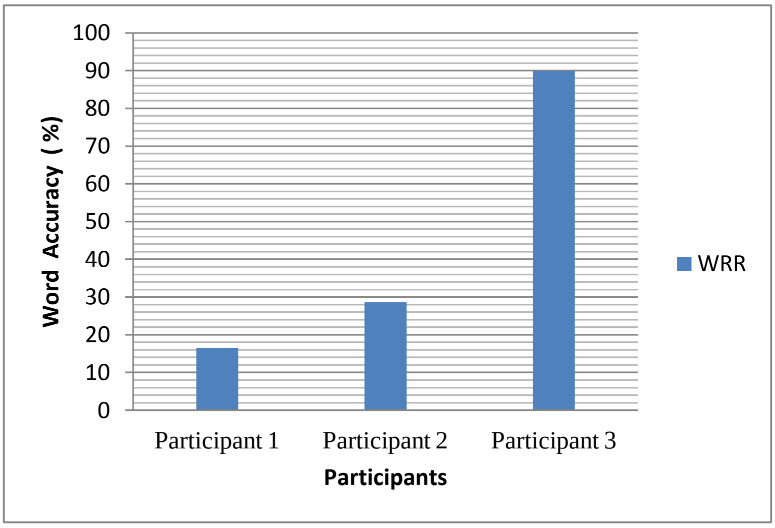

Figure 5: Pilot Test Sample graph

3.1.1.1 Problems Identified

During the training session for NDNS, It was observed that reading the training passage from the screen was difficult for users, and also using the library environment resulted in ASR system failure to recognize any user's speech as a result of the need to speak quietly. To solve the first problem: all the available training passages for NDNS were printed out so users could read conveniently without squinting their eyes or slouching over the PC to read. The second problem identified was avoided by using other environments other than the library. It was also observed that there was the need for a second ASR system in order to verify and compare the performance of one system with the other. The result of the pilot test prompted a further need to observe more participants. The experience gathered during this period was used to finalize the study design approach used for this study.

3.1.2 Criteria for Participants selection.

Selection factors are:

1. Language Group
2. Level of Education
3. Computer Literacy
4. Good Usage of the English language for learning and communication
5. Relevance of ASR system to a participant's profession or the possible possession of Windows ASR system.

The selection was planned with the aim of reaching the different South African ethnic groups who fall into the different professions (see table 6 below). After identifying an individual as belonging to one of the language groups, other selection factors mentioned above were then considered before recruiting the individual for participation. The relevance of the ASR system to an individual's profession was also considered. For instance, an ASR system will not be as relevant to an unskilled factory worker's performance or output as it will be to that of a student or a writer. Good use of the English language was required so as to reduce the effect of wrong pronunciation on the result of the evaluation. The Pre-study Questionnaire (see Appendix 1) results informed on participants' possession of the selection criteria mentioned above and these attributes formed the minimum requirement for participation. The result also informed on different accent groups used in this study. Forty individuals were approached as shown in the table 6 below, but at the commencement of the field observation, only a total of twenty participants eventually made commitments to participate.

Table 6: Participants Selection plan

Language Group	Number of Participants approached	Actual number of Participants	Comments
Afrikaans Speaking	10	6	Five of these participants claimed to have British accent, while the sixth participant claimed Afrikaans accent
Indians	5	1	Indian accent but claimed to have European influence.
Black South Africans	13	2	Both claimed to have African accent with European influence.
Migrants from other African Countries	12	11	Two claimed British accent, three others have Europe influenced accent, while the remaining six claimed to have various indigenous African accents

3.2 Study Design Approach.

The study design aimed at carrying out the following tasks in participants' natural environments: offices, homes. Within-group design was used for the task assignment. In any study involving evaluation, participants are usually assigned tasks to perform. Between- group [47] is a design approach in which the participants are divided into different groups and each group is unexposed to the treatment or task of the other groups. Sometimes the different groups are exposed to the same treatments/tasks but under different conditions in order to measure and compare the effect of the condition on performance. In this approach, the participants might be divided into control test group and a study test group .Within-group design [47] applies the same condition, the same treatments/tasks to all the participants and then compares performance. Ultimately both design approaches are set to make performance comparison. The advantage of between-groups design is that participants are not over-burdened with several tasks and so, they are not exposed to the risk of physical or mental fatigue. But the pitfall of this approach is that it could be expensive, time consuming and demands the recruitment of a large number of participants. Within-groups design has the tendency to cause fatigue, and participants' knowledge of a previous task could influence their performance on subsequent tasks but this approach provides a more economical means of carrying out an evaluation process, and it allows for optimum observation of any individual participant. The choice of within-group design for this study was impelled because of the limited number of participants available at the commencement of the study, as only twenty participants eventually took part in this study. The downside of this approach was minimized by encouraging participants to perform multiple tasks simultaneously (see tasks below). The UEMs of interest to this study are: field observation and the use of questionnaires. Field observation (observing users) was used for this study because it is a means of extracting the usability competence of a system by recording real-time performance of the system with individual users. Questionnaires are also used because the questionnaires' outcome indicates the views of the users of the system (in terms of flexibility, learnability and robustness). The questionnaires will help indicate the users' satisfaction with the ASR systems.

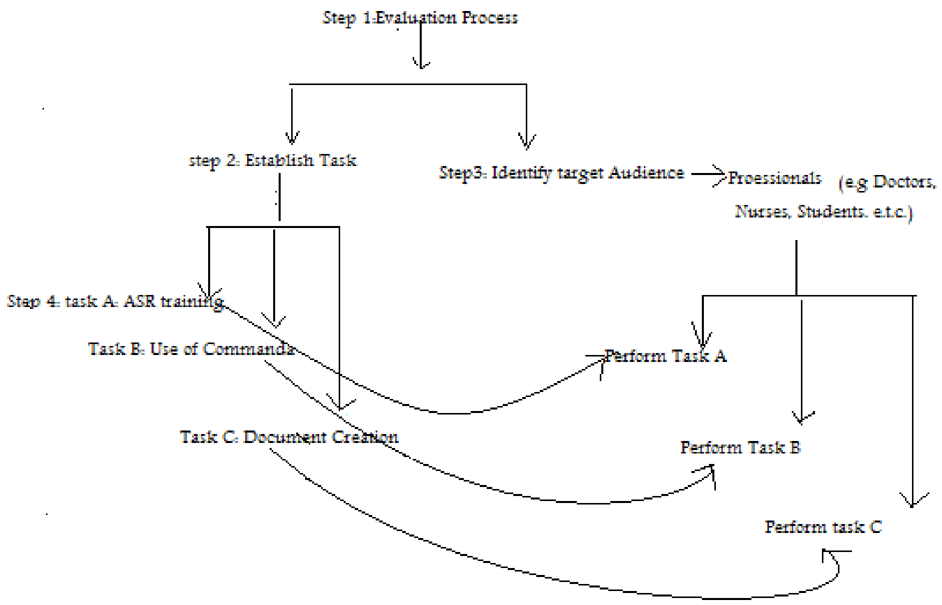

Figure 6: Evaluation Process

The field observation is structured as shown in figure 6 above.

Step 1: involved identifying the evaluation process necessary for the study.

Step 2: involved identifying and establishing task to be carried out.

Step 3: involved identifying the target audience and

Step 4: linked the tasks to the audience identified in step 3.

The tasks are:

Task A: ASR training by each of the twenty participants and for the two ASR systems to be used.
NDNS: Each participant was advised to read a training text for fifteen minutes.
WSR: Participants were advised to follow the training 'wizard' and all the participants used an average of six minutes for training WSR.

Task B: Voice commands were given to WSR so as to measure its positive response to voice commands. For this task, the ASR system was given a command to open Microsoft Word and then the Word document was used to carry out task C.

Task C: Document creation was the main task of the study and the word recognition accuracy of the created document was measured. Task C was performed simultaneously for both ASR systems in order to eliminate participants' influence on the results, and other factors such as environmental noise. (See Appendix 5 for detailed task description.)

3.3 Facilities

The study was carried out in the participants' offices, homes and classrooms. These are places where the target population of the ASR systems will naturally use their PCs. The professionals like the doctors, nurses and writers were observed using the ASR systems in their offices and the entrepreneurs were observed in their homes which doubled as their offices. Students were observed in classrooms.

3.4 Systems Requirements.

The systems acquired for this study are two commercially available ASR systems:

- NDNS version 10 (DVR edition) and
- WSR (a feature of the Windows Vista edition of PCs).

The features of the NDNS are limited as it is only for transcription of recorded audio files while Windows Vista WSR has the ability to create a document, listen to and obey voice commands, and can be used to give commands while browsing the internet.

According to the manufacturer [15], NDNS supports:

- Microsoft Windows7, 32-bit and 64-bit
- Microsoft Windows Vista SP1 and SP2, 32-bit and 64-bit
- Microsoft Windows XP SP2 and SP3, 32-bit only

And it requires:

- CPU: minimum 1 GHz Intel Pentium or equivalent AMD processor
- Processor Cache: minimum 512 KB.
- Free hard disk space: 2.5GB
- RAM: minimum 1 GB for Windows XP and Windows Vista, 2GB for Windows 7 and Windows Server 2003/2008
- Nuance-approved noise-cancelling headset microphone
- Internet connection is required for activation of NDNS.

3.5 Procedure

After the approval of the study proposal and the results of the pilot test were analyzed, the study was conducted. The evaluation procedure was conducted as follows:

At the start of the process, participants were asked to enter their first name into a register and used the serial number assigned to their names as their identification number (participant's number) throughout the study. This was to ensure that no participant's personal detail was revealed, and also to ensure that data were not mismatched. (See Appendix 4 for participants register).

Participants were given a consent form which clearly described the evaluation procedures and the roles of participants (see Appendix 3 for consent form used).

Participants were given a pre-study questionnaire to answer, to ensure that they met the requirements for participation; data of participants not meeting the requirements was not used in the result (see Appendix 1 for Questionnaire 1).

Task A: For NDNS, every participant was required to create a user profile created with the instructions below:

STEP 1: Creating new user:

- 1a: Click on the 'new' button for new users.
- 1b: Enter your new user name as 'participant #' with the number (#) given to you by the study investigator
- 1c: Select the language that best suits your English accent from the language drop bar. Then click on the 'next' button.

STEP 2: Speech Recording and Training

- 2a: Before you click on the 'next' button on new user wizard, record your speech by using the voice recorder provided, and read the passage provided for 15 minutes.
- 2b: Connect the voice recorder to the computer. Click on the 'next' button for the next two window pages of the new user wizard until you see 'browse', click on 'browse'
- 2c: After identifying your speech by listening, let the ASR system adapt to your speech.
- 2d: Take a break of 10-15 minutes.

STEP 3: Task completion

- 3a: Read to record a short passage provided by the study investigator
- 3b: Connect the voice recorder to the computer and click on transcribe on the Dragon page.
- 3c: Let the study investigator do performance/accuracy measurement.

Participants were advised to read aloud a training passage for fifteen minutes. These training passages are part of the NDNS software package. It is mandatory to train the NDNS prior to user profile set up. This passage reading was done into a voice recorder, which was then plugged into the

PC via a USB cable for voice adaptation. Participants then followed the instructions stated above (see Appendix 5 for full NDNS Task Description) for user profile set-up. Note that recording of the main task in step 3 was executed simultaneously for both ASR systems.

After the NDNS training and user profile set-up had been achieved, WSR Training was done by following the training 'wizard' prompts.

Task B: For the task of voice command and document creation, WSR was given the voice command thus: 'Start Microsoft office word 2007'. After a positive response was achieved, a short passage was dictated out to WSR while simultaneously recording the same short passage for NDNS voice recorder (Philips voice Tracer). The number of attempted command prompts before recording a positive response was also collected as data for command response measure. In this measure, the lower the number of command counts, the better the performance.

Task C: The document created by WSR was saved with users' participation number with MSV added to differentiate it from NDNS. WSR was closed and then NDNS was opened. Participants chose their participants' number and plugged in the voice recorder for transcription. The transcribed document was saved using participants' numbers with 'DNS' added to the document name.

Both WSR and NDNS documents were then printed out and the word accuracy for each system was then calculated; the number of correctly recognized words were counted and averaged over the total number of words in the short passage. This gave recognition accuracy of each system for every participant.

Users were given a post-test questionnaire to answer in order to survey user satisfaction (see Appendix 2 for Questionnaire 2).

(Note: at no given time were the two ASR systems simultaneously opened on the same PC as it is a requirement for the functionality of the NDNS application)

3.6 Data Analysis

Research Question 1: How well do the two systems under study cope with diverse accents? Data collected was analysed by calculating the recognition accuracy rate of each participant for the two ASR systems. This was done by counting the number of words correctly recognized by the system averaged over the total number of words of the passage read. The number of attempts made by a user before achieving a positive response to a command was also collected as data for the measure of command response.

Research Question 2: Will users find the two ASR systems learnable, flexible and robust? Data collected via the questionnaires were analysed to evaluate learnability, flexibility, and robustness. System performance was also compared based on their recognition accuracy rate. Users' satisfaction was also analysed. This research question is required in order to establish the usability of the ASR systems.

CHAPTER 4: RESULTS

4.1 Result Analysis

The collected data for this study are: Word accuracy rate, Command response rate, and users' satisfaction via questionnaire.

4.1.1 Data entry

Table 7 below represented the result of the recognition accuracy measured in percentage and the command response rate. The data entry was also accompanied by the profession and the language/accent group each participant belong to.

Table 7: Data Entry for Participants (see Appendix 6 for word accuracy calculation & Command response rate)

Participant #	English Accent	Profession/level of Education	WSR Word Accuracy (%)	NDNS ASR Word Accuracy (%)	WSR Command Response Rate
Participant 1	African Accent(Ghanaian)	Entrepreneur (Psychologist)	13	20	4
Participant 2	African Accent (Nigerian)	Post-graduate Student (Geology)	20	29	2
Participant 3	British Accent (SA Afrikaans)	Entrepreneur	50	90	1
Participant 4	African Accent (Nigerian)	Medical Practitioner	21	63	3
Participant 5	African Accent (Cameroonian)	Graduate (English Language)	40	54	4
Participant 6	African Accent (Nigerian)	Entrepreneur (Law graduate)	34	43	4

Participant 7	African Accent (Ghanaian)	Economist	21	15	4
Participant 8	British Accent (SA Afrikaans)	Writer (Editor for Bible-language translation)	56	85	1
Participant 9	British Accent (SA Afrikaans)	Writer	65	79	1
Participant 10	British Accent(African) (West African)	Library Science Graduate	23	71	3
Participant 11	African (Xhosa)	Undergraduate (English Language)	43	50	4
Participant 12	African (Xhosa)	Student	46	52	2
Participant 13	Afrikaans Accent	Undergraduate	63	80	1
Participant 14	African Accent (Zimbabwean-Sohan)	Undergraduate (Accountancy)	36	71	1
Participant 15	African Accent (Zimbabwean-Sohan)	Accountant	22	46	2
Participant 16	Indian Accent	Professional Nurse	06	68	2
Participant 17	Afrikaans Accent(Coloured SA)	Professional Nurse	02	13	4
Participant 18	British Accent	Medical Practitioner	50	82	1
Participant 19	African Accent	Medical Practitioner	45	66	1

Participant 20	African Accent	IT Specialist	20	48	1

Table 7 above shows the list of all the participants and it is presented in the order of participant registration. Of the twenty Participants, seven claimed to have a British accent. They claimed to have been born in Britain but migrated to South Africa, and the word accuracy for this group of participants was high. The second best word recognition belongs to six participants. These were Indians or Africans who stated that their accents had British, German, or Dutch influences because they once lived or studied in these countries. The least on the word recognition accuracy rate are participants with no influence of any European languages and such claimed having lived major parts of their lives in Africa. Results indicated the variation in recognition accuracy as a result of diverse users' accents. Table 8 below is a summary of the number of participants identified with each accent group.

Table 8: Accent grouping for participants

Accent Groups	Number of participants
British Accents	7
Indians/Africans with European-influenced accents	6
Indigenous African Accents	7

4.1.2 NDNS Data Analysis.

The data for NDNS users are given in table 9 below with the data grouped according to the English accent of the participants, and each accent group result was represented in descending order.

Table 9: NDNS Data Grouping

Accent Grouping	Participant #	Word Accuracy (NDNS in %)
British/European Accents	Participant 3 Participant 8 Participant 18	90 85 82

	Participant 13	80
	Participant 9	79
	Participant 10	71
	Participant 14	71
African/Indian Accents (With influence of European Accents)	Participant 16	68
	Participant 19	66
	Participant 4	63
	Participant 5	54
	Participant 12	52
	Participant 11	50
Indigenous African Accents	Participant 20	48
	Participant 15	46
	Participant 6	43
	Participant 2	29
	Participant 1	20
	Participant 7	15
	Participant 17	13

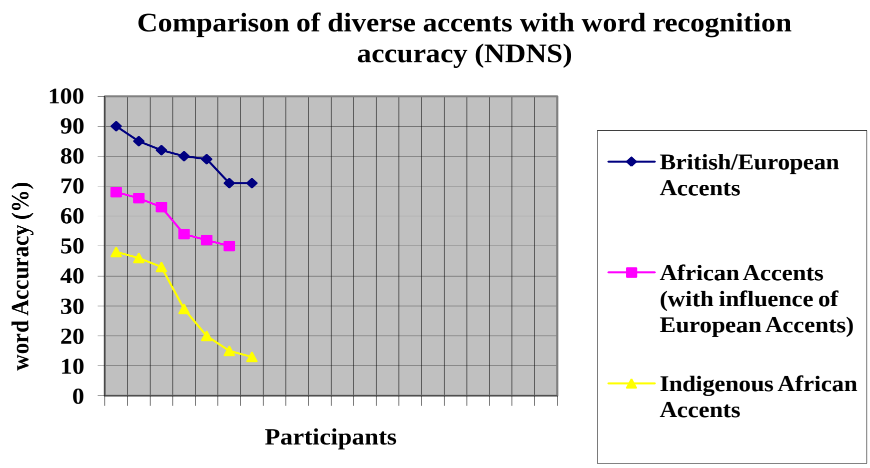

Figure 7: NDNS participant grouping

For NDNS (see figure 7 above), the recognition accuracy indicated that Participants with British accents achieved higher accuracy. This supports the assumption that the ASR system was designed for such users. While those with African accents with European influences also achieved a better result, indigenous African accents achieved the least recognition accuracy. Thus the result (see Table 9) indicated that the usefulness of ASR for African users is minimal in terms of users' productivity. During the user set-up in task A, the British accent was identified as one of the language groups covered by the ASR system. This explains the robust recognition as compared to the other language groups. This study result compared favourably with other research work on accent and its effect on ASR system recognition [5, 8, 54, 68]. These research works informed on the adverse effect of diverse accents on the recognition of ASR systems.

One line of work by Brink and Botha [8] not only confirmed that variant accents have negative impact on ASR system recognition but further compared the various South African English accents; those of native speakers (English as first language) and non-native speakers (those with mother tongues such as Zulu, Xhosa). The result of the variation was then used to enhance the language model of an existing ASR system in order to create a system which is more robust to diverse accents. Tjalve [68] also reported that accent variation has an impact on the performance of ASR system, and he identified the root cause of this effect as the inadequacy of language group dictionary (pronunciation dictionary) and the difference in the accent rules of every language group. He proposed designing an individual pronunciation dictionary which could then be integrated into a standard language model. Like the work in [68], the work of Pedersen and Diederich [54] also acknowledged that accent that differs from that of an ASR system language model results in poor recognition. Biadsy's line of work [5] focussed on designing an individual pronunciation dictionary and examined the effectiveness of this approach to ASR systems, this work also indicated that recognition rate and error occurrence depend on the similarity between the user accent and ASR language model.

The majority of errors observed in the participants’ passages were either omission or substitution errors. Omission or substitution errors occur when the system’s language model discards (omits) or wrongly transcribes a user’s speech to unrelated words or phrases. Thorough assessment of each participant’s passage indicated some common wrongly recognized words/ phrases. These words/ phrases indicated no particular pattern of occurrence in relation to each language group or any profession. Two of the most common errors are given in table 10 below:

Table 10: Comparison of original words to transcribed words

Original word/phrase:	Participants, P#-NDNS/WSR Transcription/Recognition	Original word/phrase	Participants, P#-NDNS/WSR Transcription
Typically:	P1- omitted/omitted P2-minus/omitted P3-omitted/omitted P4-omitted/omitted P5-‘that kylie’/serving P6-incarnate/omitted P7-omitted/seek P8-‘that the key’/omitted P9-italy/omitted P10-typically/omitted P11-omitted/omitted P12-omitted/omitted P13-omitted/omitted P14-typically/omitted P15-typical/omitted P16-‘thats many’/omitted P17-omitted/omitted P18-typically/separately P19-omitted/omitted P20-omitted/omitted	Various:	P1-omitted/omitted P2-omitted/omitted P3-omitted /omitted P4-omitted/omitted P5-omitted/omitted P6-omitted/omitted P7-omitted/omitted P8-theories/theories P9-areas/theories P10-various/ ‘men as’ P11-omitted/omitted P12-previously/omitted P13-omitted/omitted P14-finest/omitted P15-virus/omitted P16-history/omitted P17-omitted/omitted P18-various/periods P19-omitted/omitted P20-omitted/omitted

In Table 11 below, the sample mean for word accuracy indicated the mean recognition accuracy rate for each accent group. This also indicated low recognition accuracy for indigenous African accents. The sample means for both NDNS and WSR word accuracy were given side by side (to enable ease of comparison), alongside the mean command response of WSR.

Table 11: Sample Means for NDNS and WSR word accuracy

Accent Grouping.	Sample Mean for NDNS Word accuracy	Sample Mean for WSR word accuracy	Sample Mean for WSR and NDNS	Sample Mean for WSR Command Response.
British/European Accents	79.7	49	64.4	1.3
African Accents With influence of European Accents	58.8	33.5	46	2.7
Indigenous African Accents	30.6	18.9	24.7	2.7
Average of the whole group	56.4	33.8	45	2.2

4.1.3 WSR Data Analysis.

The data in table 12 below are those of WSR users and also grouped according to their language groups and the result was presented in descending order. The number of attempt of command response was also given.

Table 12: WSR Data Grouping

Accent Grouping	Participant #	Word Accuracy (WSR in %)	WSR Command Response Rate (# of Attempts)
British/European Accents	Participant 9 Participant 13 Participant 8	65 63 56	1 1 1

	Participant 3	50	1
	Participant 18	50	1
	Participant 14	36	1
	Participant 10	23	3
African Accents With influence of European Accents	Participant 12	46	2
	Participant 19	45	1
	Participant 11	43	4
	Participant 5	40	4
	Participant 4	21	3
	Participant 16	06	2
Indigenous African Accents	Participant 6	34	4
	Participant 15	22	2
	Participant 7	21	4
	Participant 2	20	2
	Participant 20	20	1
	Participant 1	13	4
	Participant 17	02	4

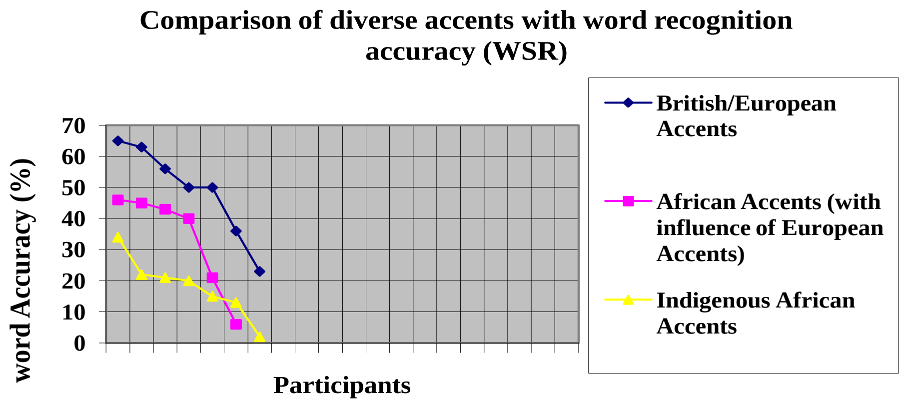

Figure 8: WSR participants grouping

For WSR (see table 11 above), the recognition accuracy indicated that participants with British accents achieved the highest accuracy with an average of 49% (though lower as compared to NDNS). This was as expected since the system is designed for such users. Much lower accuracies, averaging at 33.5 %, were achieved by the African accents with European influence. And worst of all were the indigenous African accents, with an average of 18.9 %. In figure 8 above, the results obtained in WSR also indicated the same pattern as NDNS recognition and recognition error. This result also compared favourably with the research papers [5, 8, 54, 68] discussed under NDNS result review. The words wrongly recognized are similar to those of NDNS (see table 10 above).

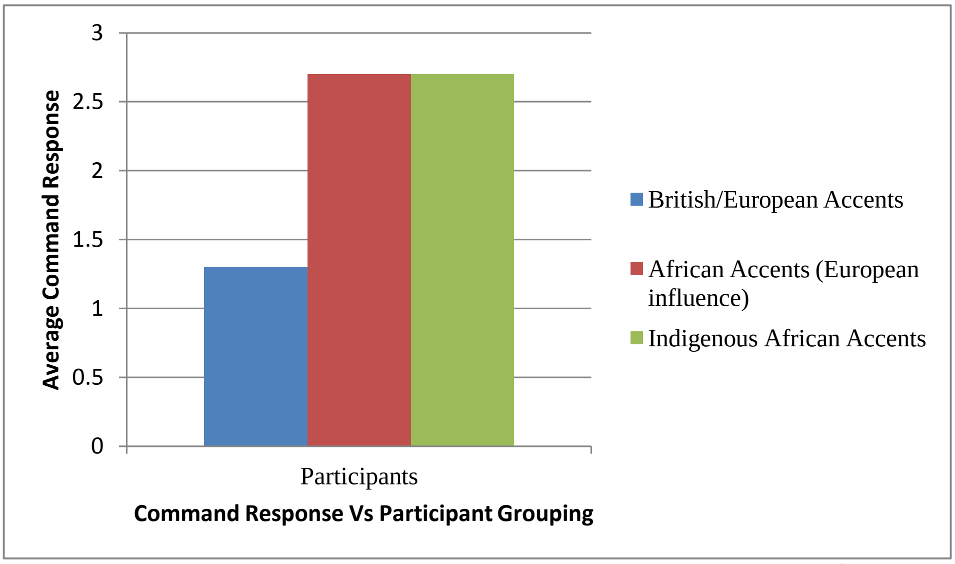

Figure 9: WSR Command Response (Mean Graph)

The command accuracy of WSR is measured by counting the number of times a user speaks a command before the command is recognised correctly and the desired response is achieved. Figure 9 illustrates the mean command response for each accent group. The command response accuracy showed an average response time of 1.3 for British/European Accents group, while the other two groups each had an average of 2.7 response time. The command response rate when compared to dictation showed a more positive outcome as regarding user performance and this can be partly attributed to the following factors:

- Commands are short sentences or phrases of words and so they are much easier and less stressful for participants to read out. The language model of WSR command feature functions as an isolated word recognizer and it is equipped with a limited vocabulary. It always presents all the possible phrases similar to the uttered command, thereby giving the user the instant opportunity to make corrections; this leads to robustness of the interface over a period of usage. Isolated word speech recognition is much more robust to accents because there are only a few possibilities for a given utterance, which reduces the probability of incorrectly identifying a command.

- Commands are precise and concise as compared to dictating several paragraphs of a passage, and thus participants are more conscious of the content and therefore more focussed.

- The WSR Command feature requires immediate feedback and interaction with the users, as users can only progress if a positive response is received from the IS. Therefore when an error occurs, users are compelled to correct this error before progressing with the task at hand. This type of interaction enhances the response rate as multiple corrections lead to ASR system frequently adapting to users speech. It adapts to any given user irrespective of the accent. However, it might have to be given the commands more than once, and ultimately it always results in correct recognition. This command attribute of WSR makes it a more usable facility to any given Windows OS user.

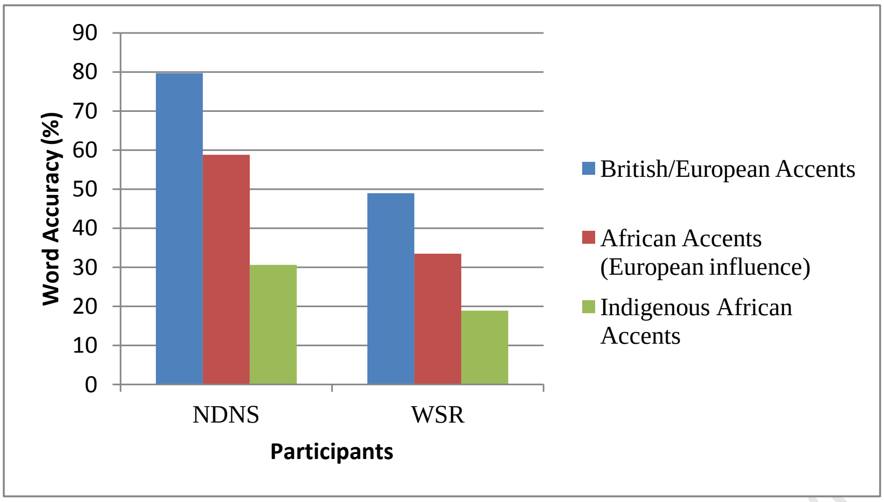

Figure 10: Comparison of NDNS and WSR Recognition accuracy (Mean Sample plot)

From the mean samples graph (see figure 10), in terms of word accuracy, the two ASR systems indicated the same pattern; which is the effect of diverse accent groups. NDNS has higher word accuracy; this shows that NDNS is more robust to multiple accents and is more adaptable to diverse accents. WSR is less robust to diverse accents as indicated by its low word accuracy rate. Participants with European accents showed higher recognition accuracy (for both NDNS & WSR) as shown in Figure 8. This result could be due in part to the fact that NDNS users did the adaptation training for 15 minutes as compared to WSR users who did the adaptation training for about 4 to 6 minutes. This result is in agreement with Hypotheses 1:

I. The average recognition rate for both NDNS/WSR will be less than 99% for all participants.
II. NDNS will achieve higher recognition accuracy; thus making it more robust compared to WSR.

4.2 Post-test Questionnaires Analysis.

The participants' opinion towards the two systems was also gathered by means of a post-study questionnaire, using the Likert-type scale, and five points used were coded thus: 1- Strongly Agree. 2- Agree. 3- Indifferent. 4- Disagree. 5- Strongly Disagree. The graph below (figure 11) compares participants' opinion of the two ASRs.

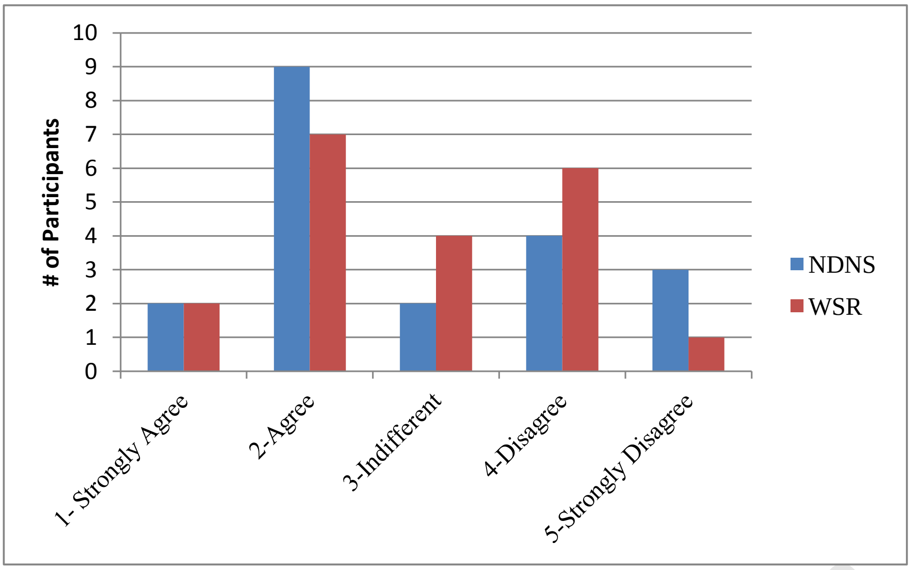

Figure 11: Comparison of the two systems for Statement 1: "This application is easy to use"

Figure 11 shows participants' response to statement 1: "This application is easy to use". Participants were advised to base this statement on the task outcome. The results indicated that more users rated NDNS higher than WSR; about 45% of participants agreed that NDNS was easy to use when compared to WSR for which only 35% of participants agreed that it was easy to use. This response implied that NDNS is more robust (see discussion on robustness in Chapter 2) to diverse accents. About 10% of both NDNS and WSR users strongly agreed with statement 1. 20% of WSR users and 10% of NDNS users were indifferent, while 30% of WSR users and 20% of NDNS users disagreed. 5% of WSR users and 15% of NDNS strongly disagreed with this statement. The overall analysis of this statement indicated that 55% of the users preferred NDNS compared to only 45% who preferred WSR. This outcome was based on performance achieved by users. This result is in agreement with hypothesis 1-II, which states that: "NDNS will achieve higher recognition accuracy; thus making it more robust compared to WSR."

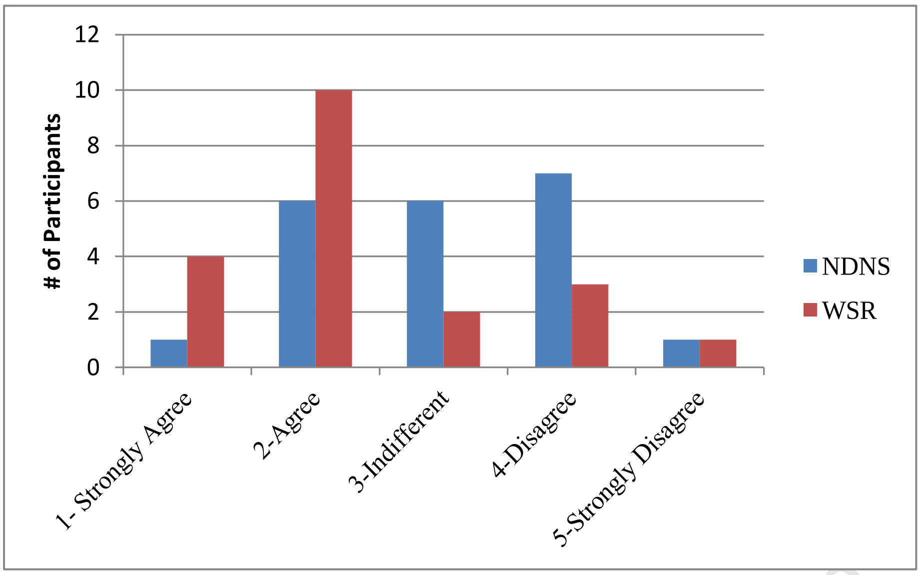

Figure 12: Comparison of the two ASR systems for Statement 2: "Creating my User's profile was easy'

The statement 2 result (see Figure 12 above) indicated that 50% of the users found WSR more learnable and flexible (see discussion on flexibility and learnability in Chapter 2), when compared to only 30% of users who agreed that NDNS was easy to create a profile for. This response compared favourably with hypothesis 2-I. About 35% of NDNS users found it difficult to create their user profile as when compared to only 15% of WSR users. About 70% of WSR users in total were pleased while using WSR for profile creation; this is in agreement with hypothesis 2-I.

Hypotheses 2:

I. Users will find WSR more learnable and flexible compared to NDNS

II. Users' satisfaction will be adversely affected by low speech recognition accuracy.

WSR system received more positive response for the rating of statement 2, and this could be due to the fact that users are used to the Windows OS screen and MS word processors. It could also be due to the average of 6 minutes spent in its training. NDNS profile setup required many Windows screens and required 30 minutes for the whole training process. The NDNS word processor is in the form of a notepad screen, and many users could find it complex, and the fact that they are using a new form of UI could have influenced the low positive response to statement 2.

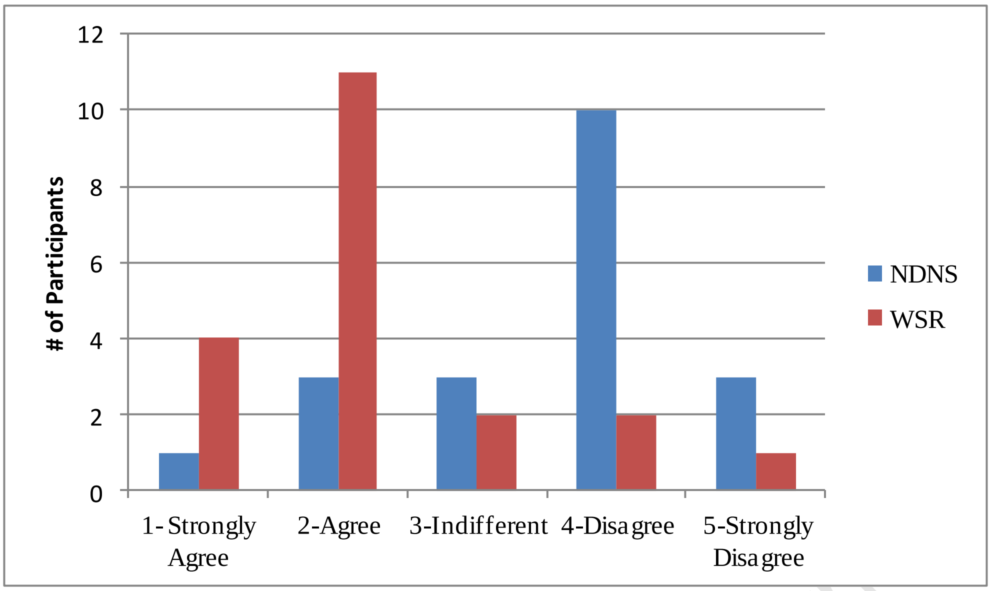

Figure 13: Comparison of the two ASR systems for Statement 3: "Training the application for my voice recognition was an easy task"

This statement was set in order to verify the outcome of statement 2. From Figure 13 above, 55% of users agreed that WSR was easier in terms of training as compared to only 15% of NDNS users. This result could be due to the fact that WSR users read the training text for about 4 to 6 minutes as compared to NDNS users who read for15 minutes. This statement outcome compared favourably with the outcome of statement 2.This statement 3 result also supported Hypothesis 2-I that WSR is more learnable and flexible.

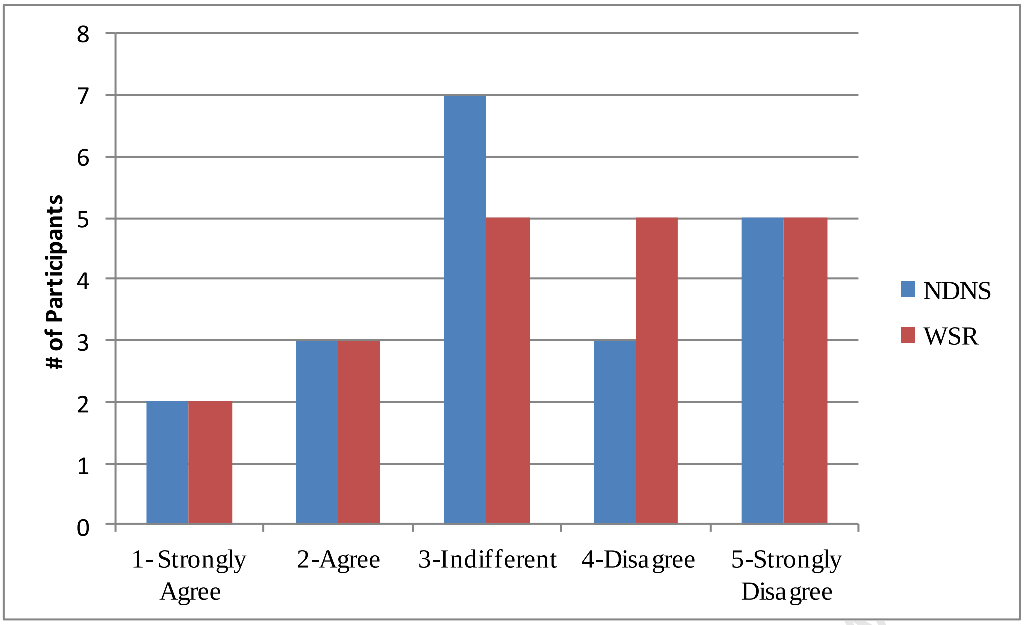

Figure 14: Comparison of the two ASR systems for Statement 4: "I find dictating to a computer easier compared to typing via keyboard."

In Figure 14, 25% of NDNS and 15 % of WSR users disagreed that dictating to a computer is easier than typing via keyboard. For both NDNS and WSR, only 10% strongly agreed to this statement and 15% agreed to this statement. This result could be due to users' familiarity with keyboards. In the study of [32], Hugunin and Zue argued that ASR systems could be easier to use if they are properly implemented into the context of use. But many of these participants were still trying to adjust to the change from the traditional keyboard as this was their 'comfort zone'. This implied that it will require a longer and regular usage of the ASR systems before a clear choice could be made.

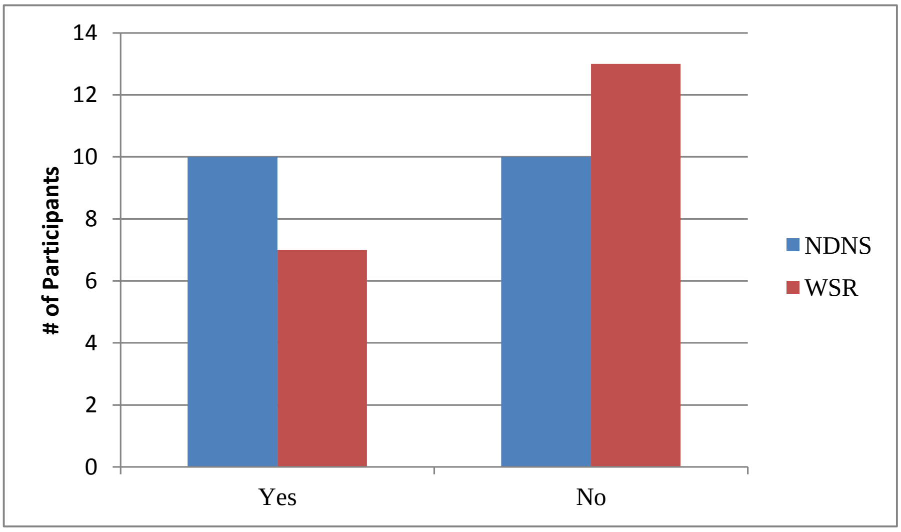

Figure 15: Comparison for the two ASR systems for Question 5: "Did you achieve your set goal while using this application?'

Figure 15 indicates users' satisfaction in terms of their task completion/achievement while using the systems. Users who achieved higher results of 60% and above tended to be more satisfied as compared to users with less than 50%. 50% of NDNS users agreed to have achieved their set goals and therefore satisfied, all of those who answered 'yes' for NDNS were found to have recorded recognition accuracy in the range of 60% to 90%. Only 35% of the WSR users agreed to have achieved their set goals and were satisfied using the WSR System. WSR users with recognition accuracy in the range of 50% to 70% were those who responded 'yes' to WSR. This statement's outcome supported Hypothesis 2-II: "Users' satisfaction will be adversely affected by low speech recognition accuracy."

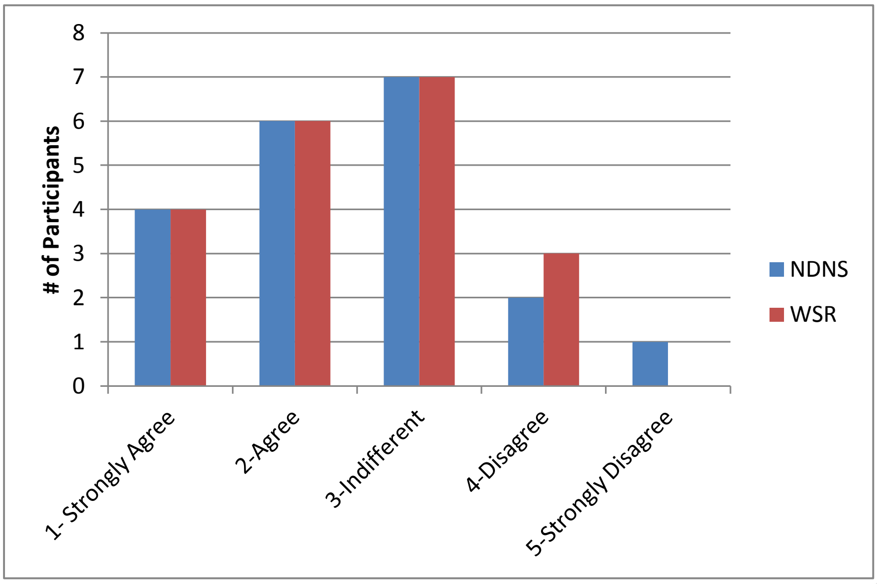

Figure 16: Comparison of the two ASR systems for Statement 6: "Using Speech Recognition is faster compared to keyboard typing."

The purpose of Statement 6 was to try and gauge user's perception of the ASR technology, and to see whether they were aware of the speed benefit. Majority of the participants' response was indifferent regarding statement 6 as illustrated in Figure 16 above. This may be due to the fact that the two ASR systems were used just once which was not sufficient to draw a stronger conclusion. It was desired during the study to encourage participants to type via keyboard and then compare the speed to that of dictating to an ASR system but the time allocated to the experiment would be exceeded and this could overburden the participants. But the work of Hugunin and Zue [32] compared favourably with this statement; they carried out an experiment on conventional keyboard/mouse and ASR systems. The ASR system was found to record 50% faster work output; they further argued that if an ASR system is properly designed for the context of use, it will provide a better result than the traditional keyboard/mouse.

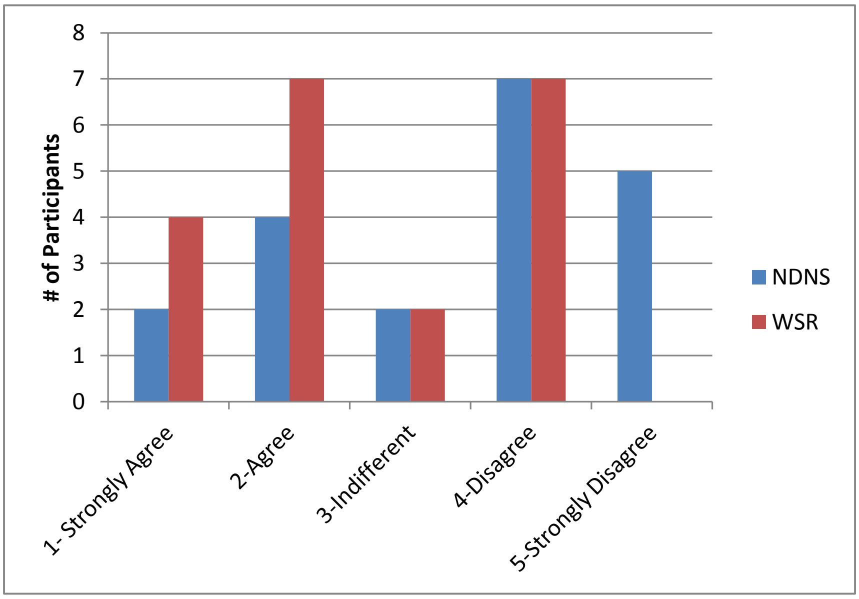

Figure 17: Comparison of the two ASR systems for Statement 7: "This application can help improve my productivity"

In Figure 17, about 35% of WSR users agreed that the system could help improve their performance compared to 20% of NDNS users; this could be due to the fact that WSR is part of the Windows operating system while NDNS is a separate application and more expensive to acquire. 25% of NDNS users strongly disagreed that it would improve their performance while no WSR users strongly disagreed with this statement.

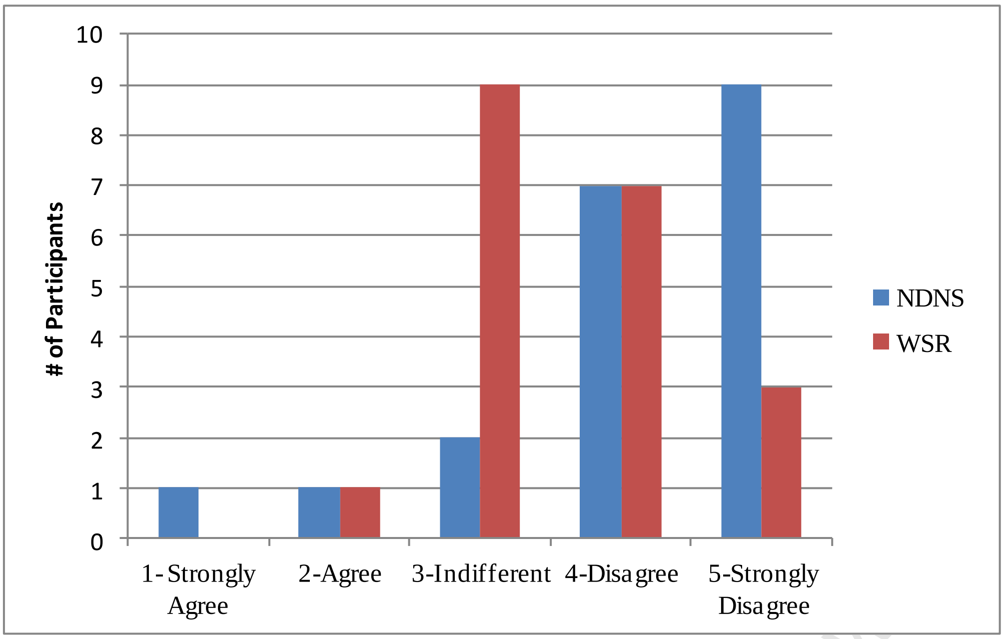

Figure 18: Comparison of the two ASR systems for Statement 8: "This application's procedures/steps are straight forward."

About 80% of both NDNS and WSR users responded negatively to this statement (from Figure 18 above). This could be due to training processes involved before getting to use the two ASR systems. Users are familiar with UIs which require little or no training prior to usage. ASR systems are new forms of UIs requiring more than the norms of a UI. In order for ASR systems to be more acceptable by users they have to have usability quality components [76]. This statement outcome showed that 80% of the users are not satisfied with the procedures of the ASR systems. It is therefore important that ASR systems evolve from the complicated training requirement as time is of utmost importance to many professionals. Organisations would also need to implement a two-way training program: cost of the time spent to train an ASR system is one, and training the personnel to use the ASR system is also at a cost.

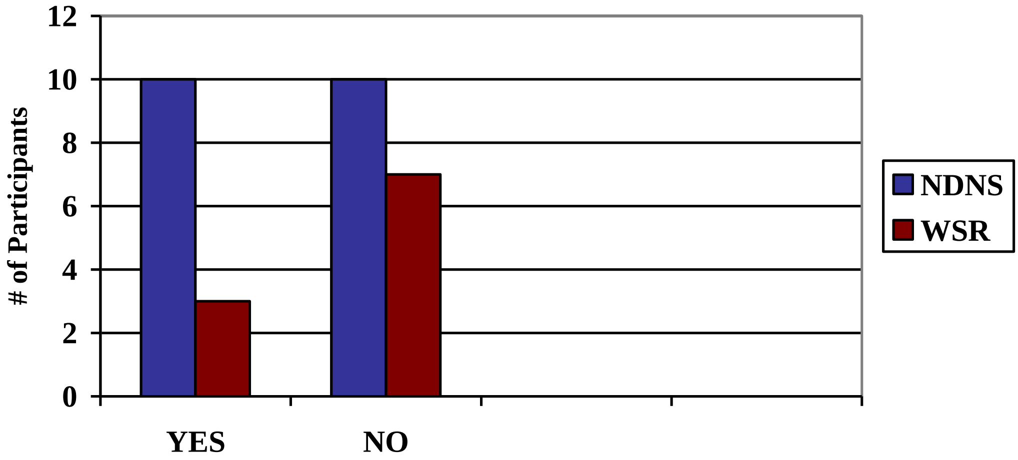

Figure 19: Comparison of the two ASR systems for Question 9: "Are you satisfied using this new application?"

According to data analysis in Figure 19; for NDNS, 50% of participants gave positive response, while the other 50% gave a negative response, it could be assumed that this result reflected the usability divide among the diverse accent groups. On the other hand 35% of WSR users gave a negative response and only 15% gave a positive response, and the rest of WSR users did not respond to this question. This statement's outcome is in agreement with the outcome of question 5 which states: 'Did you achieve your set goal while using this application?' And it also supported hypothesis 2-II which states: 'Users' satisfaction will be adversely affected by low speech recognition accuracy.'

CHAPTER 5: DISCUSSIONS AND CONCLUSIONS

5.1 Discussions.

The results obtained for both ASR systems indicated that there was variation in recognition accuracy as a result of diverse users' accents. Users of the same language group or similar accents had a word accuracy range that was similar. This led to a partial conclusion that diverse accents could influence the performance (recognition accuracy) of the ASR systems. This conclusion was based on novice users of ASR systems and within a short time frame, but in order to establish the complete behaviour of ASR system, users could be observed frequently over a longer period of time.

5.1.1 Hypothesis 1-I: The average recognition rate for both NDNS/WSR will be less than 99% for all participants.

This study result is in agreement with hypothesis 1: I. The overall word recognition accuracy for both systems was less than 90%, (see figure 8). An average of 45% was recorded for the two ASR systems for all participants. The overall result for the two systems across the different accent groups are: the British accent group recorded an average of 64%, while the African accent (with influence of European accents) recorded an average of 46%, and lastly the indigenous African accent group recorded the lowest accuracy of 25% (see table 12).

For NDNS, 65% of the participants recorded above 50% word accuracy but less than 99% accuracy. The highest word accuracy of 90% was recorded for participant 3, an entrepreneur with a British accent, and the average word accuracy for this accent group was 80%. The African/Indian accent group (with influence of European accents) recorded a maximum of 68%. This result was obtained by participant 16 (an Indian, a professional nurse), while an average of 59% was recorded for this group. The indigenous African accent recorded a mean of 31% and with a high of 48% obtained by participant 20 (an African, IT specialist).

For WSR, 25% of the participants recorded word accuracy of above 50% but less than 70%. The British accent group recorded an average of 49% and a high of 65%. The African accent group (with influence of European accent) had an average of 34% and an individual high of 46%. Lastly, the indigenous African accent group had a mean of 19% and an individual best of 34%. The result of Figure 8 confirmed that NDNS and WSR recorded accuracies less than 99%.

This result not only supported hypothesis1-I and identified performance divide along the accent groups but also showed that the 99% accuracy claim by NDNS manufacturer could not be reached by diverse accent groups in their respective environment in which ASR system would be used. The argument of Coy and Barker [11] also supported the outcome of this study, as they argued that manufacturers used 'clean speech' under controlled environments in rating ASR system performance. The manufacturer's ideal conditions are different from the normal conditions under which users are likely to use the ASR systems. This study outcome confirmed that given a normal, noisy environment and diverse accents, ASR system performance degrades significantly.

5.1.2 Hypothesis 1-II: NDNS will achieve higher recognition accuracy; thus making it more robust compared to WSR.

For NDNS, the highest word accuracy obtained was 90% when compared to WSR with 65%. These study results compared favourably with the ASR system performance reviews from Internet sources [16, 17, 18, 44, 67, 70, 71, 72, 77, 78, 81]. NDNS was rated as the best commercially available ASR system by these Internet sources. NDNS recorded an average recognition accuracy of 57% for all participants, when compared to WSR system which recorded an average of 34%. The variation in performance could be due in part to NDNS having a greater robustness (see chapter 2 for Robustness) to diverse accents. Secondly, the reason for NDNS recording a better accuracy could be due to its detailed training process, when compared to about an average of 6 minutes spent in training WSR. Comparing the best accent group performance for the two systems showed that the 7 participants with British accent recorded a better result with NDNS. These results were obtained after participants simultaneously read a passage for both ASR systems. Therefore it was not that NDNS's better performance or WSR's poor performance was due to a change in participants' mind-set. The microphone for NDNS could have helped its better performance as it could be placed close to the user's mouth, unlike WSR for which an in-built microphone was used (see Section 5.1.2.1 for the reasons why two different microphones were used). Table 13 illustrates the comparison for both systems with regard to British accent group.

Table 13: Comparison of British accent Performance (WSR and NDNS)

British/European Accents	NDNS	WSR
Participant 3	90	50
Participant 8	85	56
Participant 18	82	50
Participant 13	80	63
Participant 9	79	65
Participant 10	71	23
Participant 14	71	36

The reason for NDNS better performance could be due to other factors such as its robustness to noise, or the underlying technology implemented in its design. Furui [21] pointed out that for an ASR systems to be useful and usable, it must be robust to human factors such as: speech variation and change in environmental conditions. NDNS seems to be more robust to speech variation with regard to diverse accents and the different environments of the participants.

5.1.2.1 Technology Limitations and their Impact on the Methodology

There were limitations in the technology that influenced the methodology of this study as regards the use of two different microphones. In the analysis of the hypothesis 1-II, one of the probable reasons for NDNS higher recognition accuracy was thought to be the microphone used. This situation could not be helped (as explained below). However, it was ensured that both systems were exposed to the same speech by requiring participants to dictate to the two systems simultaneously. There were limitations which were encountered before resolving to carry out the study as described in the methodology. There are facts about these two systems which created a challenge to the best possible means of ensuring that they are exposed to the same acoustic data; these are:

Fact 1: The only thing common to both systems is that they are both ASR systems; otherwise, they are completely different. NDNS is designed to recognise recorded audio data and WSR is only designed to recognise 'live' speech and not recorded audio. Furthermore, WSR is not technologically ready for processing and recognising recorded audio data [45], as it is not a transcribing system which NDNS is even though both are ASR systems. Also, it was not possible to use one microphone while dictating simultaneously. Dictating simultaneously was preferred over using the same microphone at different times. This is because, in the latter case, it was assumed that the change in environmental conditions and in the style of speaking of the participants would affect the acoustic data more. This assumption, however, was not tested in this study and is left to future work. Therefore this study methodology treated them independent of one another by using their respective microphones.

Fact 2: It was important that the integrity of the acoustic data was assured by keeping the necessary conditions constants: ASR systems are affected by conditions such as: air, environmental noise, user behaviour, and the propagation medium such as a microphone. The user behaviour consists of several other constraints which were elaborately discussed in Section 2.3.1. In order to minimize the effect of all these conditions, the following two options were considered:

1. One microphone could be used in order to eliminate variation in acoustic data due to different channels. Although this was desirable, it meant dictating at different times, which in turn meant that environmental conditions and user behaviour could not be guaranteed to remain unchanged. It was therefore important to the discourse of this study to find a balance between all the conditions mentioned above. Option 1 favoured using the same microphone at the expense of the other conditions. Furthermore, a second drawback of this option emerged in the additional experiment described below. The external microphone performed poorly for WSR compared to the in-built microphone. This challenge in option 1 led to the consideration of option 2.

2. To consider multiple conditions by choosing to use each system with its respective microphone. This option enabled each system to be simultaneously used with its respective microphone under the same air, environmental noise and user behaviour. This option was thought to likely minimize (by keeping the conditions constant) the effect of external conditions that might negatively impact these systems. Therefore, this study chose to use both systems simultaneously. By so doing, the same environmental conditions and user behaviour would apply for both ASR systems. Furthermore, it was thought that systems perform best with the equipment for which they were designed. An experiment (described below) was conducted which supports this point of view, at least in the case of WSR.

In the original study, fact 1 was observed. In order to take into account the impact of microphone used on the results of WSR accuracy when comparing it to NDNS accuracy, another small study was conducted. In this study, five participants were recruited according to the original study methodology, (although there was no consideration for the participants' accent).Training and set-up of WSR was also conducted as described in the methodology. The focus of this second observation was to observe recognition rates of WSR for the external microphone (which had been used for NDNS) as compared with those for the computer's own in-built microphone. Each participant dictated a passage with the use of the in-built microphone and then dictated the same passage the second time using the external microphone. The result obtained is shown in table 15 below: using WSR with the in-built microphone was represented as WSRM while using WSR with the external, Nuance-approved microphone was represented as NDNSM.

Table 14 below shows the result obtained and this result will be tagged 'Result-WSR', to minimize the possibility of mixing up this result with the result of the original study.

The sample mean recorded for WSRM was approximately 27%, while that of NDNSM was 16%. WSR original study recorded a sample mean of 34%; note that the original study was compiled over twenty participants as opposed to five participants in result-WSR. The result-WSR obtained showed that the sample mean for WSRM was higher than the sample mean of NDNSM by 11.02%.The average command response for NDNSM was four times the average of WSRM. This result-WSR, though very enlightening, showed a pattern which would not have been desirable in the original study, as, if the NDNS microphone had been used without this knowledge, it would have been impossible to have identified its possible negative impact on the recognition accuracy and ultimately the study. Furthermore, this study avoided this likely pit-fall by assuming that it was safe to use systems with their factory-fitted or recommended accessories. It was assumed that it would be safe to use each system respective microphone.

Table 14: Result-WSR Data Entry

WSR Participants (WSR P#)	Word accuracy for in-built Microphone. (WSRM) in %	Word Accuracy for external Microphone (NDNSM) in %	No of commands (WSRM)	No of Commands (NDNSM)
WSRP1	45.93	39.25	2	3
WSRP2	17.28	14.81	1	8
WSRP3	35.80	1.73	1	4
WSRP4	17.78	9.14	1	10
WSRP5	15.76	12.52	2	4
Sample Mean	26.51	15.49	1.4	5.8
Original study Sample mean	33.8	-	2.2	-

.

This result-WSR also gave more insight on the behaviour of ASR systems and the effect of using various accessories can have on performance. However, it is important to note that the original study did focus on user accent, and the need for two systems was necessary in order to complement one another. The usability divide recorded among the twenty participants could only have happened as a result of a common feature in the different user accents. NDNS recorded a usability divide pattern and this same pattern was also recorded in WSR as represented in figure 10 above, despite each ASR system being independent of the other. Since all participants were made to undergo the same condition, it will suffice to conclude that the methodology did not create the usability divide recorded in the findings. This second observation showed that the methodology adapted was sufficient to make the partial conclusions reached in the original study. Finally, this second experiment requires future work as its aims differ completely from the research objectives of the main study.

From figure 20 below, the average command also showed that a great deal of effort was put in by participants before recording a positive response for NDNSM. This result as illustrated in figure 20 showed that on the average, participants needed to give four times more commands than for WSRM

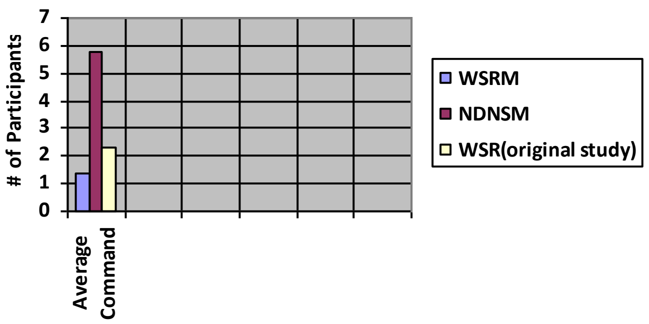

Figure 20: NDNSM/WSRM Command response Sample Mean

in order to record a positive response. For instance, a particular participant received a single positive response only after repeating the command up to ten times (see table 14 above). Looking back at the result recorded in the original study, it showed that NDNSM was almost three times worse than WSR (original study) as shown in figure 20 above.

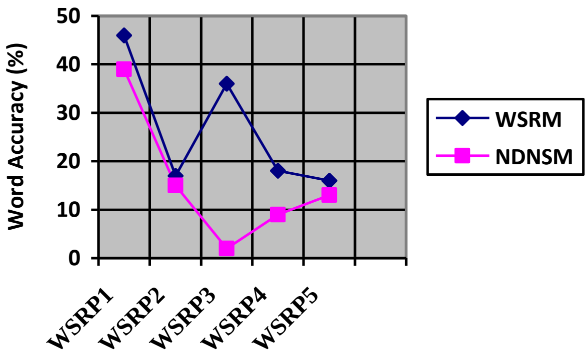

Figure 21: NDNSM/WSRM Recognition accuracy

In figure 21 above, the result-WSR showed that despite using the Nuance-approved microphone, WSR recognition accuracy was not better than the result of the WSR in-built microphone or that of NDNS. This showed that there is more to NDNS system that made it more robust than WSR. WSRM recorded an individual best of 46% and the lowest was 16%, while NDNSM recorded an individual best of 39% and with the lowest record of 2%. In partial conclusion, this second study also confirmed that NDNS is more robust than WSR irrespective of the microphone used, therefore supporting hypothesis1-II.

5.1.3 Hypothesis 2-I: Users will find WSR more learnable and flexible compared to NDNS

The majority of the response of statements 2: "Creating my User's profile was easy", and statement 3: "Training the application for my voice recognition was an easy task" in the post-study questionnaire are in support of hypothesis 2-I. This hypothesis is supported by the outcomes of statements 2 and 3 as discussed earlier. Even though NDNS was more accurate, participants were not familiar with its features .The importance of usability quality components cannot be overemphasised as they are the lifeline of a good system, and irrespective of the cultural differences, every user desires to encounter a flexible and learnable system.

5.1.4 Hypothesis 2-II: Users' satisfaction will be adversely affected by low speech recognition accuracy.

The majority of users who responded positively to statement 5 which states: "Did you achieve your set goal while using this application?" and statement 9 that states: "Are you satisfied using this new application?" were found to have recorded a recognition accuracy of above 50%. The results showed that users are likely to be satisfied when they achieve their set goals while using any system. Nielsen [76] pointed out the importance of ensuring optimal performance of a system. He argued that the output of an organisation depends on the usefulness/usability of the system implemented. He further advised on the need to implement a system that will benefit not a few but all parties involved; as this will reflect on the organisational productivity. The result of this study presented an overall performance of less than 50%. If organisations in which these participants belong to were to require the use of existing ASR systems, overall productivity might be adversely affected. These study results mean that performance will vary in any given context of application of the ASR system. It will rather be more beneficial if organisations can acquire ASR systems specifically designed for the different accent groups in South Africa; acquiring such ASR systems will be more user specific and therefore likely to improve output.

5.2 Conclusion

This study was limited in the number of participants and the amount of time they spent on the experiment. Therefore a follow up study is needed in order to confirm these preliminary findings. Conclusions reached in this study identified the existence of a correlation between diverse accents and ASR systems' recognition accuracy. In order to make a more concrete conclusion on the performance of the ASR systems, a further study into the behaviour of ASR systems will be required.

This study has shown there is a need for research into ASR systems for African users. In the early days of the invention of the computing machine, it was easier for the developing world to acquire such innovative design, as the design was universally adaptable. From the understanding of inclusive design [33], it could be said that the ASR system used in this study was not inclusively designed, and that it was specifically designed for some language groups. In order for Africans to keep up with new innovations such as ASR, there is need to adapt such innovations to meet Africa's uniqueness and needs.

The continuous research and development in ASR technology has given rise to its widespread usage in gadgets with user interfaces (UIs) [62], but such implementation is only as good as its performance in any given context of use. It will be to the advantage of users in Africa to have an ASR system that truly represents their uniqueness and diversity. At the research level, pronunciation models could be designed for the different accent groups in South Africa and then incorporated into existing ASR systems extensively. This could result in more universally useful systems within the South African society. There are recent works such as those of Biadsy [5] and Pedersen et al [54] who have designed pronunciation models and incorporated them into existing ASR systems and recorded a significant improvement in the performance of the ASR systems. In conclusion, future work should carry this further and establish: (1) that usability evaluation is expedient for the success of any technology and (2) that there is a need to research and develop an ASR for Africans by Africans (for English, which facilitates communication among diverse language groups in many African countries), and (3) the need to build ASR systems for African languages.

5.2.1 Challenges.

During this study, the most challenging aspect was the ability to reach the right participants; participants needed to be diverse in language group. This then required one-on-one requests and an informal discussion with all the participants. This led to visiting hospitals, police stations, libraries and offices. From the institutions visited, the library and hospital yielded a positive response while response/ permission from Western Province Police HQ to use police officers is still pending as at the time of this write-up. UCT students were not solely used as this would not be a true representation of the target population. It was really challenging to always meet the participants at their own convenient time, as this then required the experimenter to cancel all other schedules in order to keep the appointments. Disappointments of many participants making appointments and not keeping to the set date and time also contributed to the limitations of the results obtained. The financial commitment to reach all participants at the scheduled time also contributed to the challenges faced.

From a technical perspective, there was a challenge in comparing the two systems under study. The two systems accept data in different ways; hence the data processed was unique to each system and it is likely not identical. In future work, a means to provide identical data will be important in order to verify the results of the comparisons made here. It must, however, be noted that the comparison between the two systems was secondary to the main aim of the study, which was to compare the effect of accents within each system separately. The observed trends were found to be the same in both systems. This observation, that African accents were greatly disadvantaged under both WSR and NDNS suggests that future research needs to focus on this group.

5.3 Recommendations.

From the experience gathered, it is recommended that a further study be done to include as many participants as possible and also the possibility of designing an ASR system for Africans by Africans should be considered. Manufacturers could make ASR systems that are more useful and usable by ensuring that the pronunciation models suit the target populations' accents.

References

1. Abran, A., Khelif, A. Suryn, W., and Seffah, A. 2003. Usability Meanings and Interpretations in ISO Standards. In *Software Quality Journal, 11.*Kluwer Academic Publishers, the Netherlands. pp 323-336.

2. Agaram, K.K., Keckler, S.W., Burger, D. Characterizing the SPHINX Speech Recognition. Tech Report: TR200-33. Computer Architecture & Technology Laboratory, Department of Computer Science, the University of Texas at Austin. Available from: http://www.research.ibm.com/acas/projects/00agaram.pdf

3. Automatic Speech Recognition. Available from www.babelfish.org/asr-free.htm

4. Barnard, E., Davel, M. H. and Van-huyssteen, G.B. Speech Technology for information access: a South African case study. Human Language Technologies Research Group Meraka Institute CSIR South Africa. Available from http://www.meraka.org.za/pubs/barnard10speechtechnology.pdf

5. Biadsy, F.2011. Automatic dialect and accent recognition and its application to speech recognition. Doctorate dissertation. Graduate school of Arts and Sciences, Columbia University.

6. Bogren, L., Fallman, D., Henje, C. 2009. User-centred Inclusive Design: Making public Transport accessible. Umea Institute of Design, Umea University, Sweden.

7. Bridle, J., Deng, L., Picone, J., Richards, H. B., Ma, J., Kamm, T., Schuster, M., Pike, S., and Regan, R. 1998. An investigation of segmental hidden dynamic models of speech coarticulation for automatic speech recognition. In the workshop on *Language Engineering, Center for Language & speech processing*. The John Hopkins University.

8. Brink J.D., Botha E.C. An acoustic comparison of the vowels and dipthongs of first and second language south African English. Department of Electronics and Computer Engineering, University of Pretoria,002,South Africa.

9. Cairns, P. and Cox, A. L. 2008. *Research methods for human- computer interaction.* Cambridge University Press, NY, USA.

10. CMU Sphinx Development Available from http://cmusphinx.sourceforge.net/

11. Coy, A. and Barker, J. 2007. An automatic speech recognition system based on the scene analysis account of auditory perception. In *Speech Communication*. 49(5), Elsevier Science Publishers, B.V. Amsterdam, the Netherlands. [doi: 10.1016/j.specom.2006.11.002].pp 384-401.

12. Clarkson, P.J., Keates, S., Dowland, R. 1999. The design and control of assistive devices. In *Proceedings of the International on Engineering Design (ICED)'99.* 1 pp425-428.

13. Digital Divide: The three stages. Available from http://www.useit.com/alertbox/digital-divide.html

14. Doe, H. L.1998. Evaluating the effect of Automatic speech Recognition Word Accuracy. Masters dissertation. Faculty of the Virginia Polytechnic Institute and State. Retrieved September 10, 2010 from: http://scholar.lib.vt.edu/theses/available/etd-7598-165040/unrestricted/thesis1.pdf

15. Dragon NaturallySpeaking Professional. Available from http://www.nuance.com/naturallyspeaking/products/editions/professional.asp

16. Dragon NaturallySpeaking 8 vs. IBM viavoice Pro USB Edition Version 10. Available from: http://www.axistive.com/dragon-naturallyspeaking-8-vs-ibm-viavoice-pro-usb-edition-version-10.html

17. Dragon NaturallySpeaking 10 Preferred review & rating by PCMAG. Available from http://www.pcmag.com/article2/0,2817,2327350,00.asp

18. Dragon NaturallySpeaking review by Trusted Reviews. Available from http://www.trustedreviews.com/Nuance-Dragon-NaturallySpeaking-10-Preferred-Wireless_Software_review

19. English Pronunciation, available from www.studyenglishtoday.net/english-phonetics.html

20. English speakers, available from www.countrystudies.us/south-africa/50.htm

21. Furui, S.1995. Towards the ultimate synthesis/recognition systems. In *proceedings of the National Academy of science*, 92. pp 10040-10045.

22. Glass, J. 2003.MIT- Introduction to Automatic Speech Recognition Available from http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-345-automatic-speech-recognition-spring-2003/lecture-notes/lecture1.pdf

23. Google mobile search. Available from: www.google.com/mobile/google-mobile-app/

24. Gough D.H. English in South Africa. :http://www.ru.ac.za/media/Gough_article.pdf

25. Halverson, C. A., Horn, D. B., Karat, C., and Karat, J. 1999.The beauty of errors: Pattern of error correction in desktop speech systems. In *Human-computer interaction, INTERACT '99,* Sasse, M. A. and Johnson, C., Eds. IOS Press, the Netherlands.

26. Hamidi, F., Spalteholz, L. and Livingston, N. 2008.Web -speak: A customizable Speech-based web navigation Interface for people with disabilities. In *The 23rd Annual International Technology &Persons with Disabilities Conference.* Available from http://www.cse.yorku.ca/~fhamidi/resources/WebSpeak.pdf

27. Hamill, M., Young, V., Boger, J., and Mihaailidis, A. 2009. Development of an automated speech recognition interface for personal emergency response systems. In *Journal of NeuroEngineering and Rehabilitation,6*.The Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, ON, Canada. Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2713250/

28. Hanna, N., Ayer, D. J., Ridnour, R.E., and Gordon, G.L. 1995. New Product development practices in consumer versus business product organizations. In *Journal of Product & Brand Management*, 4(1), pp33-55.[doi: 10.1108/10610429510083749].

29. Harper and Row .Navigator: An Apple Concept from 1987. Available from http://www.brainpickings.org

30. Heiskanen, E. and Hyvonen, K. 2006. Consumer Involvement in developing services based on speech recognition. *In proceedings of 19th Bled eConference eValues*. National Consumer Research Centre Finland. Available from http://ecom.fov.uni-mb.si/proceedings.nsf/0/c9451de85a994d6ac125718000335b31/$FILE/39_Heiskanen.pdf

31. How language works, available from http://www.indiana.edu/~hlw/PhonProcess/

32. Hugunin, J. and Zue, V. 1997. On the design of effective speech based interfaces for desktop applications. Spoken language systems group, Laboratory for computer science, Massachusetts Institute of technology Cambridge, Massachusetts 02139 USA. Available from http://groups.csail.mit.edu/sls/publications/1997/eurospeech97-hugunin.pdf

33. Inclusive Design. Available from: http://www.tiresias.org/research/guidelines/inclusive.htm

34. Juang, B.H. and Rabiner, L. R. 2004. Automatic Speech Recognition – A Brief History of the Technology Development. Retrieved from Elsevier Encyclopaedia of Language and Linguistics.

35. .Karat, J. Karat, C.M. 2003. The evolution of user-centered focus in the human-computer interaction field. In *IBM systems Journal*, 42, pp532-541.

36. Kaushik, M., Trinkle, M., and Hashemi-sakhtsari, A.2010. Automatic Detection and Removal of Disfluencies from spontaneous speech. School of Electrical & Electronic Engineering, The University of /Adelaide, Adelaide, South Australia. Available from: http://assta.org/sst/SST-10/SST2010/PDF/AUTHOR/ST100011.PDF

37. Keates, S., Clarkson, P.J., and Harrison L.2000. Towards a practical Inclusive design approach. Department of Engineering, University of Cambridge. UK.

38. Kirchhoff, K. and Bilmes, J. A. 1999.Statistical Acoustic Indications of Coarticulation. In *14th int. Cong. of phonetic sciences.* Applied Computer science Group, University of

Bielefeld, Germany. ICSI/U.C. Berkeley, Dept. Of EECS, CS Division, Berkeley, USA. Available from: http://docs.google.com/viewer

39. Lamine, K., Nganga, J., and Odiwanor, E. Principles that support usability flexibility. Available from http://docs.google.com/viewer?

40. Learnability as a Usability Method: What is learnability? Available from : http:// www.interface-design-tool.com/learnability-as-a-usability-method

41. Lee, C. A. Efficiency in Cockpit: A comparison of keypad-entry and voice recognition systems. Available from: http://www.carrielee.net/pdfs/Cockpit.pdf

42. Lee, K. 1989. *Automatic Speech Recognition: the Development of the Sphinx System*. Kluwer Academic Publishers, 3300 Dordrecht, the Netherlands.

43. Lewis, C., Polson, P., Wharton, C., and Rieman, J. 1990. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *proceedings ACM CHI'90 Conference.* pp 235-242.

44. List of best speech recognition software. Available from http://www.mytopdozen.com/Best_Speech_Recognition_Software.html

45. Microsoft Vista, available from: www.microsoft.com/enable/.../windowsvista/speech.aspx http://answers.microsoft.com/en-us/windows/forum/windows_7-sound/using-speech-recognition-in-windows-7-to/3695bae0-8ae6-4e5d-b557-c445c034abc5

46. Monk, A.F. 2002. Fun, Communication and dependability: extending the concept of usability. Closing plenary paper at HCI2002. Available from http://www.users.york.ac.uk/~am1/moukHCI02.pdf

47. Newman, W. M. and Lamming, M.G. 1995.*Interactive System Design*. Addison-Wesley Longman Publishing co, Boston, USA.

48. Nielsen, J. 1994. Heuristic evaluation. In Nielsen, J. and Mack, R. (Eds), *Usability Inspection Methods.* John Wiley & Sons, NY.25-64.

49. Nielsen, J. 1993. *Usability Engineering* .Morgan Kaufmann Publishers, San Francisco, USA.

50. Nielsen, J. 1995. Usability Inspection Methods. Available from: http://www.sigchi.org/chi95/proceedings/tutors/jn_bdy.htm

51. Norman, K.L. 2008. *Cyberpsychology: An introduction to Human- Computer interaction.* Cambridge University Press, New York.

52. Nuance Utilities. South Africa Price & Reviews available from http://www.shopbot.co.za/m/?m=nuance+dragon+naturally+Speaking+10

53. Patel, N., Agarwal, S., Rajput, N., Kumar, A., Nanavati, A. N., Dave, P., Parikh, S. T. 2008. Experiences designing a voice interface for rural India. Stanford University HCI Group, Computer Science Department, Stanford, CA 94305. http://www.stanford.edu/~neilp/pubs/slt2008.pdf

54. Pedersen C., Diederich J. Accent classification using support vector machines. School of Information Technology and Electrical Engineering, the University of Queensland, st. Lucia, Australia.

55. Rabiner, L. and Juang, B. 1993. *Fundamentals of speech recognition*. Pearson Professional Education, New Jersey, USA.

56. Ravden, S. J. & Johnson, G. I. 1989. E*valuating usability of human computer interfaces: a practical method.* Halsted Press, New York, USA.

57. Roach, P. 2001. Connected Speech and Coarticulation. *Phonetics.* pp53-61. Retrieved from: http://www.personal.rdg.ac.uk/~llsroach/phon2/asscoareli-into.htm

58. SA's PC Users base reaches 5million. Available from: http://www.southafrica.info/about/science/computers-230606.htm

59. Schneider, S. J., Cantor, D., Heller, T. H., and Brick, P.D. 2002. Pretesting Interactive voice response/automated speech recognition surveys. In *proceedings of International conference on Questionnaire development, Evaluation & testing Methods*, South Caroline Maryland. Available from: http://www.jpsm.umd.edu/qdet/final_pdf_papers/heller.pdf

60. Scholtz, J. 2004. Usability Evaluation. National Institute of Standards and Technology. Available from http://www.itl.nist.gov/iad/IADpapers/2004/Usability%20Evaluation_rev1.pdf

61. Schultz, T. and Wand, M.2009. Modelling coarticulation in EMG-based continuous speech recognition. In *speech communication Journal.* Available from http://csl.ira.uka.de/fileadmin/media/publication_files/SchultzWand-SCJ10-ModelingCoarticulation.pdf

62. Sculley J. & Byrne J.A. 1987.Odyssey: Pepsi to Apple

63. Sherwani, J., Yu, D., Paek, T., Czerwinski, M., Ju, Y.C., and Acero, A. 2007. VoicePedia: Towards Speech-based Access to Unstructured Information. Carnegie Mellon University, Pittsburgh, PA, USA. Available from http://www.cs.cmu.edu/~jsherwan/pubs/voicepedia.pdf

64. Sirpa, R. 2000. Experiences with usability evaluation methods. Doctoral dissertation. Department of Computer Science & Engineering, Helsinki University of Technology.

65. Speech Recognition review. Available from http://www.myspeech.com/

66. Threfall, K.d.1999. Using focus groups as consumer research tool. *Journal of marketing practice* 5(4).pp 102-105.

67. The New York Times: Voice recognition software review. 2008. Available from http://www.nytimes.com/2008/08/07/technology/personaltech/07pogue.html

68. Tjalve M. 2007. Accent Features and Idiodictionaries: on Improving Accuracy for accented speakers in ASR. Doctorate Dissertation. Department of Phonetics and Linguistics, University College London.

69. Tomlin, W.C. Time vs. Impact on Design. Available from: http://www.usefulusability.com/wp-content/uploads/2009/11/UsabilityImpactOnDesign.JPG

70. Top Best voice or Speech Recognition software. 2010. Available from http://ceoworld.biz/ceo/2010/01/25/top-best-voice-or-speech-recognition-software

71. "Top Ten Reviews". Available from http://voice-recognition-software-review.toptenreviews.com/

72. "Top 7 Speech Recognition" Available from http://win.downloadatoz.com/tutorial/2210,top-speech-recognition-software.html

73. Usability Evaluation Methods. Retrieved from: www.usabilityhome.com

74. Usability testing. Retrieved from: www.wikipedia.org/wiki/usability_testing

75. Usability Testing. Available from: http://www.cs.cmu.edu/~cprose/LTI-6-UsabilityTesting.pdf

76. Usability 101: Introduction to Usability. Available from : http://www.useit.com/alertbox/20030825.html

77. Viavoice vs. Dragon NaturallySpeaking. Available from: http://www.associatedcontent.com/article/1607829/via_voice_vs_dragon_naturally_speaking_pg2.html?cat=15

78. Viavoice vs. Dragon NaturallySpeaking Available from http://www.rsi-relief.com/2005/04/viavoice-vs-dragon-naturally-speaking

79. Virtanen, T. 2009. Noise-Robust Automatic Speech Recognition. Tampere University of technology, Department of signal processing. Available from http://www.cs.tut.fi/courses/SGN-4106/robust_audio.pdf

80. Voice Recognition software review IBM ViaVoice 10.Available from http://www.gizmag.com/go/7484/

81. "Voice recognition software review" Available from http://www.ConsumerSearch.com/voice-recognition-software

82. Weinstein, C.J. 1995. Military & Government applications of human- machine communication by voice. In *proceedings of the National Academy of Science*, 92. pp 10011-10016.

83. What is inclusive design? Available from: http://www.sensorytrust.org.uk/information/inclusive_design.html

84. Xerox Alto image. Available from: http://en.wikipedia.org

85. Yazgan, A., and Saraclar, M. 2004. Hybrid Language models for out of vocabulary word detection in large vocabulary conversational speech recognition. In *Proceedings of int. Conf. Of Acoustics, speech, and signal processing.* AT&T Labs-Research 180 Park Avenue Florham Park, NJ. [doi:10.1109/ICASSP.2004.1326093] pp745-748.

Appendices

Appendix 1: QUESTIONAIRE PRE- STUDY

EVALUATION OF THE USABILITY OF AN AUTOMATIC SPEECH RECOGNITION SYSTEM

QUESTIONAIRE (PRE-STUDY)

ARE YOU AN AFRICAN? (IF YES, WHERE IN AFRICA)
YES ☐ NO ☐

WHAT IS YOUR FISRT LANGUAGE?

☐

DO YOU SPEAK ENGLISH?
YES ☐ NO ☐

HAVE YOU EVER USED ENGLISH LANGUAGE FOR WRITING PURPOSES?
YES ☐ NO ☐

HAVE YOU EVER USED ENGLISH LANGUAGE FOR LEARNING PURPOSES?
YES ☐ NO ☐

ARE YOU A COMPUTER LITERATE?
YES ☐ NO ☐

DO YOU USE COMPUTER FOR LEARNING PURPOSES?
YES ☐ NO ☐

DO YOU KNOW WHAT AUTOMATIC SPEECH RECOGNITION APPLICATION IS?
YES ☐ NO ☐

HAVE YOU EVER USED AN AUTOMATIC SPEECH RECOGNITION APPLICATION?
YES ☐ NO ☐

ARE YOU A MALE (M) OR A FEMALE (F)?
M ☐ F ☐

NOTE: Please ensure that you read and sign the consent form before proceeding.

Appendix 2: QUESTIONAIRE –POST STUDY

EVALUATION OF THE USABILITY OF AN AUTOMATIC SPEECH RECOGNITION SYSTEM

QUESTIONAIRE 2 (POST- STUDY)

1. THIS APPLICATION IS EASY TO LEARN.

☐ Strongly Agree. ☐ Agree ☐ Indifferent. ☐ Disagree. ☐ Strongly Disagree.

2. CREATING MY USER'S PROFILE WAS EASY:

☐ Strongly Agree. ☐ Agree. ☐ Indifferent. ☐ Disagree. ☐ Strongly Disagree.

3. TRAINING THE APPLICATION FOR MY VOICE RECOGNITION WAS AN EASY TASK?

☐ Strongly Agree. ☐ Agree. ☐ Indifferent. ☐ Disagree. ☐ Strongly Disagree.

4. I FIND DICTAING TO A COMPUTER EASIER COMPARED TO TYPING VIA KEYBOARD:

☐ Strongly Agree. ☐ Agree. ☐ Indifferent. ☐ Disagree. ☐ Strongly Disagree.

5. DID YOU ACHIEVE YOUR SET GOAL WHILE USING THIS APPLICATION?

☐ YES ☐ NO

6. USING VOICE RECOGNITION IS FASTER COMPARED TO KEYBOARD TYPING:

☐ Strongly Agree ☐ Agree ☐ Indifferent. ☐ Disagree ☐ Strongly Disagree.

7. THIS APPLICATION CAN HELP IMPROVE MY PRODUCTIVITTY:

☐ Strongly Agree. ☐ Agree. ☐ Indifferent . ☐ Disagree. ☐ Strongly Disagree.

8. THIS APPLICATION PROCEDURES/STEPS ARE STRAIGHT FORWARD:

☐ Strongly Agree. ☐ Agree. ☐ Indifferent. ☐ Disagree. ☐ Strongly Disagree.

9. ARE YOU SATISFIED USING THIS NEW APPLICATION?

☐ YES ☐ NO

Appendix 3: Consent Form

Evaluation of the Usability of an Automatic Speech

Recognition System

Consent Form

Part A: The study

The purpose of this study is to evaluate the effectiveness of an Automatic Speech Recognition system.

An Automatic Speech Recognition system is a software application which recognizes human speech and processes it as input and displays a text version of the speech as output or uses the output as commands for another application usage.

As a participant in this study, you are required to use these applications as you will use your keyboard for input purposes and the applications response/performance to your speech is recorded and evaluated.

There are no risks involved in participating in this study as you will only be using a computer, voice recorder, and your voice.

As a participant, you remain anonymous in the data collected, and your privacy will not be compromised in any form.

You are participating in this study voluntarily and at any stage of the study you can withdraw with no liabilities.

Part B: Consent of Participant

I have received a description of what the study entails and I understand that it is the application's performance that is being measured and not mine.

I understand that I am a volunteer in this study and I do not stand to benefit financially or otherwise in this study.

I am participating as a willing individual and so can ask questions; withdraw at any time during the study.

As a participant my personal information will not form any part of the data collection and I will remain anonymous throughout the study.

I give my consent to be part of this study.

Signature

Appendix 4: **Participants Entry Register**

PARTICIPANT #	NAME/OCCUPATION	NATIONALITY/FIRST LANGUAGE	ASR1 PERFORMANCE/ACCURACY(NDNS)	ASR2 PERFORMANCE/ACCURACY(WSR)	WSR RESPONSE TO COMMAND ACCURACY
PARTICIPANT 1					
PARTICIPANT 2					
PARTICIPANT 3					
PARTICIPANT 4					
PARTICIPANT 5					
PARTICIPANT 6					
PARTICIPANT 7					
PARTICIPANT 8					
PARTICIPANT 9					
PARTICIPANT 10					
PARTICIPANT 11					
PARTICIPANT 12					
PARTICIPANT 13					
PARTICIPANT 14					
PARTICIPANT 15					
PARTICIPANT 16					
PARTICIPANT 17					
PARTICIPANT 18					
PARTICIPANT 19					
PARTICIPANT 20					
PARTICIPANT 21					
PARTICIPANT 22					
PARTICIPANT 23					
PARTICIPANT 24					
PARTICIPANT 25					
PARTICIPANT 26					
PARTICIPANT 27					
PARTICIPANT 28					
PARTICIPANT 29					
PARTICIPANT 30					

Appendix 5: Task Description

EVALUATION OF THE USABILITY OF AN AUTOMATIC SPEECH RECOGNITION (ASR) SYSTEM

TASK DESCRIPTION

USING WSR: Follow the steps provided by the Windows VISTA SETUP 'WIZARD'.

USING DRAGON NATURALLY SPEAKING (NDNS):

STEP 1: Creating new user

1a: Using Dragon NaturallySpeaking, you click on: 'click here to start NaturallySpeaking'

1b: click on the 'new' button for new users.

1c: enter your new user name as 'participant #' with the number (#) given to you by the study investigator.

1d: select the language that best suit your English Accent from the language drop bar. Then click on the 'next' button.

STEP 2: Speech Recording and Training

2a: Before you click on the 'next' button on new user wizard, record your speech by using a voice recorder provided, read a passage provided for 15minutes.

2b: connect the voice recorder to the computer. Click on 'next' button for the next two window pages of the new user wizard until you see 'browse' click on 'browse'

2c: after identifying your speech by listening, let the ASR adapt to speech.

2d: take a break of 10-15mimutes.

STEP 3: Task completion

3a: read a short passage provided by using the voice recorder

3b: connect voice recorder to the computer and click on transcribe on the Dragon page.

3c: let the study investigator do performance/accuracy measurement.

Appendix 6: Word Recognition Rate Calculation

Total number of words in the passage = N_T

Number of words correctly recognized = N_c

Word Accuracy Rate = W_c

Word Accuracy Rate, $W_c = N_c / N_T * 100$

Command Response Rate = number of voice command prompts